

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 5, 2003, 16:58:44 ; Search time 50 Seconds
(without alignments)
1244.416 Million cell updates/sec

Title: US-09-688-672A-14

Perfect score: 1969

Sequence: 1 MGRARIIDPTISAIDGLYDL.....RAPVEADAGGGQKVLVRNV 392

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000.

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq 19Jun03.*

```
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1969	100.0	392	20 AAW73765	M. tuberculosis an
2	1969	100.0	392	20 AAW73555	M. tuberculosis an
3	1969	100.0	392	22 AAU08226	Polypeptide encode
4	1969	100.0	392	22 AAU01882	M. tuberculosis an
5	1969	100.0	392	23 ABU05959	M. tuberculosis an
6	1969	100.0	392	23 AA229714	Mycobacterium sp.
7	1969	100.0	392	23 AAE17578	Mycobacterium spec
8	1963	99.7	543	22 AAU01905	M. tuberculosis an
9	1947	98.9	408	22 AAU01886	M. tuberculosis an

10	1947	98.9	408	22 AAU01887	M. tuberculosis an
11	1713	87.0	788	22 AAU01903	M. tuberculosis an
12	1258	63.9	394	23 ABU05402	M. tuberculosis an
13	1174	59.6	242	22 AAU01883	M. tuberculosis an
14	1061	53.9	219	22 AAU01884	M. tuberculosis an
15	1055	53.6	744	22 AAU01902	M. tuberculosis an
16	1055	53.6	815	22 AAU01904	M. tuberculosis an
17	658	33.4	137	22 AAU01885	M. tuberculosis an
18	148.5	7.5	269	23 ABP38616	Staphylococcus epi
19	145	7.4	911	23 ABP26657	Streptococcus poly
20	140	7.1	911	23 ABP29744	Streptococcus poly
21	138.5	7.0	689	23 ABP65485	Bifidobacterium lo
22	138.5	7.0	7339	24 AAU016358	Human translocated
23	138	7.0	898	18 AAW31853	Mycobacterium tube
24	134.5	6.8	528	22 AAB82611	Spider recombinant
25	134	6.8	318	21 ABB81229	Mycobacterium tube
26	133.5	6.8	712	21 AAB08630	Amino acid sequenc
27	133.5	6.8	730	21 AAB08631	Fusion protein com
28	133	6.8	1020	23 AAE18321	Collagen like prot
29	133	6.8	1077	14 AAR37751	CLP 3.1 monomer co
30	133	6.8	1077	17 AAR93261	CLP 3.1 polymer se
31	133	6.8	1077	19 AAW57657	Collagen-like poly
32	132.5	6.7	291	22 AAB82608	Spider recombinant
33	131.5	6.7	162	15 AAR57103	Carrot glycine-ric
34	130.5	6.6	660	20 AAY01303	Human tropoelastin
35	129.5	6.6	501	16 AAR76076	Mycobacterium anti
36	129.5	6.6	515	21 AAY69135	Amino acid sequenc
37	129.5	6.6	571	21 AAY69071	Amino acid sequenc
38	129.5	6.6	646	18 AAW27178	Nephila clavipes s
39	129.5	6.6	655	23 ABB76672	Protein related to
40	129.5	6.6	681	22 AAB82609	Spider recombinant
41	129.5	6.6	691	22 AAB82610	Spider recombinant
42	129.5	6.6	698	20 AAY01302	Human tropoelastin
43	129.5	6.6	698	21 AAY69069	Amino acid sequenc
44	129.5	6.6	730	19 AAW46315	Human elastin cont
45	129.5	6.6	730	23 AAO17360	Human elastin. Ho

ALIGNMENTS

RESULT 1

AAW73765
ID AAW73765 standard; Protein; 392 AA.

AC AAW73765;

DT 24-MAR-1999 (first entry)

DE M. tuberculosis antigen clone hTcc#1 protein sequence.

KW Antigen; M. tuberculosis protein; immunotherapy; tuberculosis; diagnosis; infection.

OS Mycobacterium tuberculosis.

PN WO9853075-A2.

PD 26-NOV-1998.

PF 20-MAY-1998; 98WO-US10407.

PR 05-MAY-1998; 98US-0073010.

PR 20-MAY-1997; 97US-0859381.

PA (CORI-) CORIYA CORP.

PI Alderson MR, Campos-Neto A, Dillon DC, Skeiky YAW;

DR WPI, 1999-045314/04.

XX N-PSDB; AAX01177.

PT Polypeptide comprising immunogenic Mycobacterium tuberculosis

PT antigen - useful for immunisation against M. tuberculosis infection
 PR to treat or prevent tuberculosis, and in diagnosis of tuberculosis
 XX
 PS Claim 2; Page 87-88; 100pp; English.
 XX
 CC This sequence represents an immunogenic portion of a Mycobacterium
 CC tuberculosis antigen of the invention. The polypeptides are useful for
 CC immunotherapy to treat or prevent tuberculosis (especially in humans);
 CC e.g. they can be included with an acceptable carrier in pharmaceutical
 CC compositions or included in vaccines, and administered to induce
 CC protective immunity in a patient against M. tuberculosis. Tuberculosis is
 CC a chronic, infectious disease generally caused by M. tuberculosis
 CC infection, and if left untreated typically results in serious
 CC complications and death. Fusion proteins containing the antigen, or DNA
 CC molecules can similarly be included with an acceptable carrier in
 CC pharmaceutical compositions or in vaccines and administered as above. The
 CC polypeptides are also useful for diagnosis of tuberculosis, by contacting
 CC dermal cells with at least one polypeptide and detecting an immune
 CC response (especially induration) on the patient's skin. Inhibiting the
 CC spread of tuberculosis requires vaccination and accurate diagnosis, since
 CC antibiotic therapy may not be effective due to the existence of an
 CC asymptomatic but contagious stage and to patient non-compliance. The
 CC polypeptides overcome concerns of safety and efficacy of current
 CC vaccination with live bacteria (usually Bacillus Calmette-Guerin) and
 CC lack of sensitivity and specificity of existing diagnostic techniques.
 XX Sequence 392 AA;
 SQ

Query Match	100.0%;	Score 1969;	DB 20;	Length 392;
Best Local Similarity	100.0%;	Pred. No. 2e-163;		
Matches 392;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MSRAFIIDPTISAI	DGLDGLGIGIPNOGGTLYSSLEVF	EKALEEELAAAFPGDGWLSAA 60
Db	1	MSRAFIIDPTISAI	DGLDGLGIGIPNOGGTLYSSLEVF	EKALEEELAAAFPGDGWLSAA 60
Qy	61	DXYAGKRNHNHVF	FOELADLRQLISLIHDQANAVQTRDRI	LEGAKKGLEFVRPVAVDLT 120
Db	61	DXYAGKRNHNHVF	FOELADLRQLISLIHDQANAVQTRDRI	LEGAKKGLEFVRPVAVDLT 120
Qy	121	YIPVVGHALSAAF	QAPFCAGAMVVGALAVLVKTLINATQLL	KLLAKLAELVAAAIAID 180
Db	121	YIPVVGHALSAAF	QAPFCAGAMVVGALAVLVKTLINATQLL	KLLAKLAELVAAAIAID 180
Qy	181	IISDVADI	IKGTLGWBEFFITNALNGKLWDKLTGWVTGL	FSRGSNSNLESFFAGVPGLT 240
Db	181	IISDVADI	IKGTLGWBEFFITNALNGKLWDKLTGWVTGL	FSRGSNSNLESFFAGVPGLT 240
Qy	241	GATSGLSQV	TGILFGAAGLSASGLAHADSLASSASL	PALAGIOGGSGFGGLPSLAOVHAA 300
Db	241	GATSGLSQV	TGILFGAAGLSASGLAHADSLASSASL	PALAGIOGGSGFGGLPSLAOVHAA 300
Qy	301	STRQALRPRAD	GPVGAALAEQVGGGSQLVSAQSGOGMGP	VGVMGMHPSGASKGTITTKY 360
Db	301	STRQALRPRAD	GPVGAALAEQVGGGSQLVSAQSGOGMGP	VGVMGMHPSGASKGTITTKY 360
Qy	361	SEGAAGTDE	AEAPVEADAGGQKVLVRNVY 392	
Db	361	SEGAAGTDE	AEAPVEADAGGQKVLVRNVY 392	

RESULT 2
AAW73655
ID AAW73655 standard; Protein; 392 AA.
XX
XX AAW73655;
XX
XX
DT 24-MAR-1999 (first entry)
XX
XX M. tuberculosis antigen clone hTcc#1 protein sequence.
DE
XX
XX Antigen; M.tuberculosis protein; immunotherapy; tuberculosis; diagnosis;
KW infection.

```

Db      301  STFGALPRADGPVGAAGVGGQSLVSAQSGQNGGPGVGGWHPSSGASKGTTTKY 360
      361  SEGAAAGTDAERAPVEADAGGQKVLVRNV 392
      361  SEGAAAGTDAERAPVEADAGGQKVLVRNV 392

RESULT 3
AAU08226
ID  AAU08226 standard; Protein; 392 AA.
XX
AC  AAU08226;
XX
DT  17-DEC-2001 (first entry)
XX
DE  Polypeptide encoded by Mycobacterium tuberculosis clone H7Cch1.
XX
KW  Tuberculosis; Mycobacterium infection; gene therapy; anti bacterial;
KW  immunostimulant; clone H7CC.
XX
OS  Mycobacterium tuberculosis.
XX
PN  WO200162893-A2.
XX
PD  30-AUG-2001.
XX
PF  26-FEB-2001; 2001WO-US05992.
XX
PR  25-FEB-2000; 2000US-0185037.
PR  08-AUG-2000; 2000US-0223828.
XX
PA  (CORI-) CORIXA CORP.
XX
PI  Campos-Neto A, Skeiky Y, Ovendale P, Jen S, Lodes M;
XX
WPI: 2001-536638/59.
DR  N-PSDB; AAS12487.
XX
XX
PT  An isolated polypeptide comprising a Mycobacterium antigen, e.g., from
PT  Mycobacterium tuberculosis, useful in a vaccine for enhancing an immune
PT  response to and inhibiting development of a Mycobacterium infection -
XX
PS  Example 1; Page 156-157; 161pp; English.
XX
CC  The present invention relates to the isolation of Mycobacterium
CC  tuberculosis antigen polypeptides (e.g. Tb224) and the nucleic acids
CC  encoding them. The invention describes compounds and methods for the
CC  diagnosis of tuberculosis or for inducing protective immunity against
CC  tuberculosis. The compounds comprise at least one immunogenic portion
CC  of one or more Mycobacterium proteins and nucleic acid molecules
CC  encoding such polypeptides. The Mycobacterium proteins and nucleic acid
CC  molecules encoding them can be used in diagnostic kits for the detection
CC  of Mycobacterium infection in patients and biological samples. The
CC  compounds of the invention and antibodies directed against the
CC  Mycobacterium proteins may be used in vaccines for immunisation against
CC  Mycobacterium infections. The nucleic acids encoding the Mycobacterium
CC  proteins may be used in gene therapy. The present sequence represents
CC  the polypeptide encoded by M. tuberculosis clone H7Cch1.
XX
SQ  Sequence 392 AA;

Query Match      100.0%; Score 1969; DB 22; Length 392;
Best Local Similarity 100.0%; Pred. No. 2e-163;
Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  MSRAFIIDPTISAIDGLYDLLGIGIPNQGIIYSSLEYFEKALEELAAAFPGDGLGSA 60
Db      1  MSRAFIIDPTISAIDGLYDLLGIGIPNQGIIYSSLEYFEKALEELAAAFPGDGLGSA 60
QY      61  DKYAGKRNHNHVFQELADRLQLSLIHQDQNAVQTTDILEGAKKGLFVRPVAVDLT 120
Db      61  DKYAGKRNHNHVFQELADRLQLSLIHQDQNAVQTTDILEGAKKGLFVRPVAVDLT 120

```

```

QY      121  YIPVGHASNAFOAPFCAGAMAVVGGALAYLVVKTLLINATQILLKLAELVAAAIAD 180
Db      121  YIPVGHASNAFOAPFCAGAMAVVGGALAYLVVKTLLINATQILLKLAELVAAAIAD 180
QY      181  IISDVADIIGKTLGEVWEFITNALNGKLWDLKLTGVTGLFSSGWNLSFFPAGVFGLT 240
Db      181  IISDVADIIGKTLGEVWEFITNALNGKLWDLKLTGVTGLFSSGWNLSFFPAGVFGLT 240
QY      241  GATSGLSQVTCGLFGAAGLSASSGLAHADSLASSASLSPALAGIGGGSGFGGLPSLAQHAA 300
Db      241  GATSGLSQVTCGLFGAAGLSASSGLAHADSLASSASLSPALAGIGGGSGFGGLPSLAQHAA 300
QY      301  STQALRPRADGPVGAAGVGGQSLVSAQSGQMGPGVGMGMHPSGASKGTTTKY 360
Db      301  STQALRPRADGPVGAAGVGGQSLVSAQSGQMGPGVGMGMHPSGASKGTTTKY 360
QY      361  SEGAAAGTDAERAPVEADAGGQKVLVRNV 392
Db      361  SEGAAAGTDAERAPVEADAGGQKVLVRNV 392

RESULT 4
AAU01882
ID  AAU01882 standard; Protein; 392 AA.
XX
AC  AAU01882;
XX
DT  29-AUG-2001 (first entry)
XX
DE  M. tuberculosis antigen H7CCH1/Mtb40.
KW  HTCC#1; Mtb40; antigen; vaccine; tuberculosis; AIDS;
KW  acquired immunodeficiency disease.
XX
OS  Mycobacterium tuberculosis.
XX
FH  Key
FT  Peptide
FT  1..20
FT  /label= Peptide 1
FT  /note= "Used for T-cell epitope mapping"
FT  Peptide
FT  14..33
FT  /label= Peptide 2
FT  /note= "Used for T-cell epitope mapping"
FT  Peptide
FT  27..46
FT  /label= Peptide 3
FT  /note= "Used for T-cell epitope mapping"
FT  Peptide
FT  40..59
FT  /label= Peptide 4
FT  /note= "Used for T-cell epitope mapping"
FT  Peptide
FT  53..72
FT  /label= Peptide 5
FT  /note= "Used for T-cell epitope mapping"
FT  Peptide
FT  66..85
FT  /label= Peptide 6
FT  /note= "Used for T-cell epitope mapping"
FT  Peptide
FT  79..98
FT  /label= Peptide 7
FT  /note= "Used for T-cell epitope mapping"
FT  Peptide
FT  92..111
FT  /label= Peptide 8
FT  /note= "Used for T-cell epitope mapping"
FT  Peptide
FT  105..124
FT  /label= Peptide 9
FT  /note= "Used for T-cell epitope mapping"
FT  Peptide
FT  118..137
FT  /label= Peptide 10
FT  /note= "Used for T-cell epitope mapping"
FT  Peptide
FT  131..150
FT  /label= Peptide 11
FT  /note= "Used for T-cell epitope mapping"
FT  Peptide
FT  144..163
FT  /label= Peptide 12

```

FT Peptide /note= "Used for T-cell epitope mapping"
 FT 157..176
 FT /label= Peptide 13
 FT /note= "Used for T-cell epitope mapping"
 FT 170..189
 FT /label= Peptide 14
 FT /note= "Used for T-cell epitope mapping"
 FT 183..202
 FT /label= Peptide 15
 FT /note= "Used for T-cell epitope mapping"
 FT 196..215
 FT /label= Peptide 16
 FT /note= "Used for T-cell epitope mapping"
 FT 209..228
 FT /label= Peptide 17
 FT /note= "Used for T-cell epitope mapping"
 FT 222..241
 FT /label= Peptide 18
 FT /note= "Used for T-cell epitope mapping"
 FT 235..254
 FT /label= Peptide 19
 FT /note= "Used for T-cell epitope mapping"
 FT 248..267
 FT /label= Peptide 20
 FT /note= "Used for T-cell epitope mapping"
 FT 261..280
 FT /label= Peptide 21
 FT /note= "Used for T-cell epitope mapping"
 FT 274..293
 FT /label= Peptide 22
 FT /note= "Used for T-cell epitope mapping"
 FT 287..306
 FT /label= Peptide 23
 FT /note= "Used for T-cell epitope mapping"
 FT 300..319
 FT /label= Peptide 24
 FT /note= "Used for T-cell epitope mapping"
 FT 313..332
 FT /label= Peptide 25
 FT /note= "Used for T-cell epitope mapping"
 FT 325..344
 FT /label= Peptide 26
 FT /note= "Used for T-cell epitope mapping"
 FT 339..358
 FT /label= Peptide 27
 FT /note= "Used for T-cell epitope mapping"
 FT 352..371
 FT /label= Peptide 28
 FT /note= "Used for T-cell epitope mapping"
 FT 365..384
 FT /label= Peptide 29
 FT /note= "Used for T-cell epitope mapping"
 FT 373..392
 FT /label= Peptide 30
 FT /note= "Used for T-cell epitope mapping"

WO200124820-A1.

12-APR-2001.

10-OCT-2000; 2000WO-US28095.

07-OCT-1999; 99US-0158338.

07-OCT-1999; 99US-0158425.

(CORI-) CORIXA CORP.

Skeiky Y, Reed S, Houghton RL, McNeill PD, Dillon DC, Lodes ML;

WPI; 2001-290576/30.

N-PSDB; AAS03773.

Vaccinating against Mycobacteria infections in mammals using fusion

PT proteins comprising combinations of heterologous antigens
 XX Example 2; Fig 6; 168pp; English.
 XX The sequence represents Mycobacterium tuberculosis H37Rv (also known
 CC as H37Rv), an M. tuberculosis antigen. Compositions comprising at least
 CC 2 heterologous antigens, as a fusion protein, and vectors expressing the
 CC fusion proteins are used as vaccines to prophylactically immunise
 CC mammals (especially humans) against infection by Mycobacterium. The
 CC compositions contain at least 2 heterologous antigens that increase the
 CC serological sensitivity of individuals infected with tuberculosis, a
 CC disease frequently affecting patients with acquired immunodeficiency
 CC disease, AIDS.
 XX Sequence 392 AA;
 SQ
 Query Match 100.0%; Score 1969; DB 22; Length 392;
 Best Local Similarity 100.0%; Pred. No. 2e-163;
 Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 MSRAFIIDPTISAIDGLYDLIGIPNOGILYSSLEYFEKALEELAAAFPGDGLGSA 60
 Db 1 MSRAFIIDPTISAIDGLYDLIGIPNOGILYSSLEYFEKALEELAAAFPGDGLGSA 60
 Qy 61 DKYAGKRNHNHVFQELADLDROLISLIHQANNAVQTRDILEGAKKGLFVRPVDLT 120
 Db 61 DKYAGKRNHNHVFQELADLDROLISLIHQANNAVQTRDILEGAKKGLFVRPVDLT 120
 Qy 121 YIPVGHLSAFAQPCAGAMAVGGALAYLVVKTINATOLLKLAELVAAIAAD 180
 Db 121 YIPVGHLSAFAQPCAGAMAVGGALAYLVVKTINATOLLKLAELVAAIAAD 180
 Qy 181 IISDVADIKGTIGEYWEFITNALNGLKELWDLKTCWVTGLFSRGSNLESEFFAGVPG 240
 Db 181 IISDVADIKGTIGEYWEFITNALNGLKELWDLKTCWVTGLFSRGSNLESEFFAGVPG 240
 Qy 241 GATSGLSQVTLFGAAGLSASSGLAHADSLASSASIPALAGTGGSGFGGLPSLAQVHAA 300
 Db 241 GATSGLSQVTLFGAAGLSASSGLAHADSLASSASIPALAGTGGSGFGGLPSLAQVHAA 300
 Qy 301 STRQALRPRADGPGVGAAGAEQVGSQSVLSAQSQSGMGPGVGMGHPSSGASKGTTTKY 360
 Db 301 STRQALRPRADGPGVGAAGAEQVGSQSVLSAQSQSGMGPGVGMGHPSSGASKGTTTKY 360
 Qy 361 SEGAAAGTDAERAPVEADAGGQKVLVRNV 392
 Db 361 SEGAAAGTDAERAPVEADAGGQKVLVRNV 392
 RESULT 5
 ABU05959
 ID ABU05959 standard; Protein; 392 AA.
 XX ABU05959;
 XX 08-APR-2003 (first entry)
 XX M. tuberculosis and M. leprae marker protein #610.
 XX Mycobacteriosis; survival; virulence; protective antigen; vaccine;
 XX mycobacterial disease; tuberculosis; leprosy.
 XX Mycobacterium tuberculosis.
 XX Mycobacterium leprae.
 XX WO200274903-A2.
 XX 26-SEP-2002.
 XX 22-FEB-2002; 2002WO-1B01973.
 XX 22-FEB-2001; 2001US-270123P.
 XX

```

PA (INSP ) INST PASTEUR.
XX
XX
XX Cole S;
XX
XX WPI; 2002-759885/82.
XX
XX Identifying and selecting genes for survival or virulence of
XX Mycobacteria by a comparative genomic analysis of the sequences of
XX Mycobacterium tuberculosis and M. leprae -
XX
XX Claim 17; Page 828; 874pp; English.
XX
XX This invention relates to a novel method for identifying essential genes
XX for survival or virulence of mycobacteria species. The method comprises
XX aligning the genomic sequence of a first mycobacterium species on a
XX genomic sequence of a second mycobacterium species and selecting a
XX polynucleotide sequence that is highly conserved in both genomes with no
XX counterparts in other bacterial genomic sequences and that corresponds
XX to an essential gene for the survival or virulence of mycobacterium
XX species. The method of the invention is useful for detecting M.
XX tuberculosis or M. leprae infection. The method reduces the number of
XX potential new targets and protective antigens for new drugs and vaccine
XX compositions to treat and prevent mycobacterial diseases, particularly
XX tuberculosis and leprosy. The present sequence represents a marker
XX protein from Mycobacterium tuberculosis and Mycobacterium leprae
XX identified using the method of the invention.
XX
XX Sequence 392 AA;
XX
XX Query Match 100.0%; Score 1969; DB 23; Length 392;
XX Best Local Similarity 100.0%; Pred. No. 2e-163;
XX Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 MSRAFIIDPTISADGLYDLIGIPNQGILYSSLEYFEKALEELAAFPDGLGSA 60
XX Db 1 MSRAFIIDPTISADGLYDLIGIPNQGILYSSLEYFEKALEELAAFPDGLGSA 60
XX
XX QY 61 DKYAGKRNHVNPFQELADLDROLISLIHQANAVQTRDILEGAKKGLFVRPVAVDLT 120
XX Db 61 DKYAGKRNHVNPFQELADLDROLISLIHQANAVQTRDILEGAKKGLFVRPVAVDLT 120
XX
XX QY 121 YIPVVGHALSAAFOAPFCAGAMAVGGALAYLVVKTLLINATOLLKLAELVAATAAD 180
XX Db 121 YIPVVGHALSAAFOAPFCAGAMAVGGALAYLVVKTLLINATOLLKLAELVAATAAD 180
XX
XX QY 181 IISDVADIKGTGLGEVWEFITNALNGLKELWDKLTGWVTGLFSRGWSNLESPFAGVPGLT 240
XX Db 181 IISDVADIKGTGLGEVWEFITNALNGLKELWDKLTGWVTGLFSRGWSNLESPFAGVPGLT 240
XX
XX QY 241 GATSGLSQVTLGFGAAGLSASSGLAHADSLASSASLPALAGIGGSGFGGLPSLAQVHAA 300
XX Db 241 GATSGLSQVTLGFGAAGLSASSGLAHADSLASSASLPALAGIGGSGFGGLPSLAQVHAA 300
XX
XX QY 301 STRQALRPRADGFPVGAAGAEQVGSQSLVSAQSGQGMGPVGMGMHPSSGASKGTTTKY 360
XX Db 301 STRQALRPRADGFPVGAAGAEQVGSQSLVSAQSGQGMGPVGMGMHPSSGASKGTTTKY 360
XX
XX QY 361 SEGAAAGTDEARAPVEADAGGQKVLVRNVV 392
XX Db 361 SEGAAAGTDEARAPVEADAGGQKVLVRNVV 392
XX
XX RESULT 6
XX AAE29714
XX ID AAE29714 standard; Protein; 392 AA.
XX
XX AC AAE29714;
XX
XX XX
XX 27-JAN-2003 (first entry)
XX
XX Mycobacterium sp. HTCC#1 antigenic protein.
XX
XX Vaccine; immunity; diagnostic agent; gene therapy; HTCC#1 antigen;
XX

```

```

KW MTB40.
XX Mycobacterium sp.
XX WO200272792-A2.
XX
XX 19-SEP-2002.
XX
XX 13-MAR-2002; 2002WO-US08223.
XX
XX 13-MAR-2001; 2001US-275837P.
XX (CORI-) CORIXA CORP.
XX
XX Skeiky Y, Brannon M, Guderian J;
XX
XX WPI; 2002-759844/82.
XX N-PSDB; AAD47091.
XX
XX New recombinant nucleic acid molecule comprising a Leishmania TSA,
XX LeIF, M15 or 6H polynucleotide, useful as vaccine to elicit protective
XX immunity against pathogenic microorganisms e.g. Leishmania and
XX Mycobacterium tuberculosis
XX
XX Disclosure; Page 103-104; 155pp; English.
XX
XX The invention relates to a recombinant nucleic acid molecule encoding a
XX fusion polypeptide. The recombinant nucleic acid comprises a heterologous
XX polynucleotide sequence encoding an antigen or an antigenic fragment from
XX Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a
XX polypeptide or its fragment. The Leishmania polynucleotide is selected
XX from TSA, LeIF, M15, and 6H polynucleotides. Sequences of the invention
XX are used in methods for eliciting immune response in mammals. They are
XX useful as vaccines to elicit protective immunity against pathogenic
XX microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
XX polypeptides are used for enhancing the expression of polynucleotides,
XX as in vivo diagnostic agents and for raising antibodies in a non-human
XX animal. The invention is used in gene therapy. The present sequence is
XX Mycobacterium sp. HTCC#1 antigenic protein. HTCC#1 is also referred
XX to as MTB40.
XX
XX Sequence 392 AA;
XX
XX Query Match 100.0%; Score 1969; DB 23; Length 392;
XX Best Local Similarity 100.0%; Pred. No. 2e-163;
XX Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 MSRAFIIDPTISADGLYDLIGIPNQGILYSSLEYFEKALEELAAFPDGLGSA 60
XX Db 1 MSRAFIIDPTISADGLYDLIGIPNQGILYSSLEYFEKALEELAAFPDGLGSA 60
XX
XX QY 61 DKYAGKRNHVNPFQELADLDROLISLIHQANAVQTRDILEGAKKGLFVRPVAVDLT 120
XX Db 61 DKYAGKRNHVNPFQELADLDROLISLIHQANAVQTRDILEGAKKGLFVRPVAVDLT 120
XX
XX QY 121 YIPVVGHALSAAFOAPFCAGAMAVGGALAYLVVKTLLINATOLLKLAELVAATAAD 180
XX Db 121 YIPVVGHALSAAFOAPFCAGAMAVGGALAYLVVKTLLINATOLLKLAELVAATAAD 180
XX
XX QY 181 IISDVADIKGTGLGEVWEFITNALNGLKELWDKLTGWVTGLFSRGWSNLESPFAGVPGLT 240
XX Db 181 IISDVADIKGTGLGEVWEFITNALNGLKELWDKLTGWVTGLFSRGWSNLESPFAGVPGLT 240
XX
XX QY 241 GATSGLSQVTLGFGAAGLSASSGLAHADSLASSASLPALAGIGGSGFGGLPSLAQVHAA 300
XX Db 241 GATSGLSQVTLGFGAAGLSASSGLAHADSLASSASLPALAGIGGSGFGGLPSLAQVHAA 300
XX
XX QY 301 STRQALRPRADGFPVGAAGAEQVGSQSLVSAQSGQGMGPVGMGMHPSSGASKGTTTKY 360
XX Db 301 STRQALRPRADGFPVGAAGAEQVGSQSLVSAQSGQGMGPVGMGMHPSSGASKGTTTKY 360
XX
XX QY 361 SEGAAAGTDEARAPVEADAGGQKVLVRNVV 392

```

Db 361 SEGAAAGTDAERAPVEADAGGQKVLVRNVV 392

RESULT 7
AAEL17578
ID AAE17578 standard; Protein; 392 AA.
AC AAEL17578;
XX
XX
DT 22-APR-2002 (first entry)
XX
XX
DE Mycobacterium species MTB40 (HTCC #1) protein.
XX
XX
KW Fusion protein; antigen; serological sensitivity; immune response;
KW tuberculosis; infection; vaccine; MTB40; HTCC #1 protein.
XX
XX
OS Mycobacterium sp.
XX
XX
PN WO2001198460-A2.
XX
XX
PD 27-DEC-2001.
XX
XX
PF 20-JUN-2001; 2001WO-US19959.
XX
XX
PR 20-JUN-2000; 2000US-0597796.
PR 01-FEB-2001; 2001US-265737P.
XX
XX
PA (CORI-) CORIXA CORP.
XX
XX
PI Skeiky Y, Reed S, Alderson M;
XX
DR WPI; 2002-147798/19.
DR N-PSDB; AAD28349.
XX
XX
PT Composition comprising MTB39 antigen and MTB32A antigen from
PT Mycobacterium species, useful for eliciting immune response in a
PT subject
XX
XX
PS Claim 9; Page 120; 136pp; English.
XX
XX
CC The present invention relates to fusion proteins containing at least
CC two Mycobacterium species antigens, nucleotides encoding them and
CC compositions comprising such fusion proteins. The present invention
CC particularly relates to nucleic acids encoding fusion proteins that
CC include two or more individual M. tuberculosis antigens which increase
CC the serological sensitivity of sera from individuals infected with
CC tuberculosis and methods for their use in diagnosis, prevention and
CC treatment of tuberculosis infection. Sequences of the invention are
CC useful for eliciting an immune response in a mammal, e.g., human,
CC immunised with BCG. They are useful in the diagnosis, treatment and
CC prevention of Mycobacterium infection. The fusion proteins and the
CC polynucleotides are useful as diagnostic tools in patients infected
CC with Mycobacterium, in vitro and in vivo assays for detecting humoral
CC antibodies or cell-mediated immunity against M. tuberculosis, for the
CC diagnosis of an infection or monitoring of disease progression, as
CC immunogens to generate or elicit a protective immune response in a
CC patient and for raising anti-M. tuberculosis antibodies in a non-human
CC animal. Sequences of the invention are also used as vaccines. MTB32A
CC fusion proteins of the invention are useful as in vivo diagnostic agents
CC for intradermal skin test. The present sequence is Mycobacterium species
CC MTB40 (HTCC #1) protein.
XX
XX
SQ Sequence 392 AA;

Query Match 100.0%; Score 1969; DB 23; Length 392;
-Best Local Similarity 100.0%; Pred. No. 2e-163;
Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSRAFIIDPTISAIDGLYDLGIGINQGGILYSSLEYFEKALEELAAAFPGDGLGSA 60
DB 1 MSRAFIIDPTISAIDGLYDLGIGINQGGILYSSLEYFEKALEELAAAFPGDGLGSA 60
QY 61 DKYAGNRRHNVFFQELADLDROLISLIHQANAVQTTDILEGAKKGLFVRPVAVDLT 120

Db 61 DKYAGNRRHNVFFQELADLDROLISLIHQANAVQTTDILEGAKKGLFVRPVAVDLT 120
QY 121 YIPVGHLSAFAQAPFCAGAMAVGGALAYLVVKTLLINATOLLKLLAKLAEIVAAIAAD 180
Db 121 YIPVGHLSAFAQAPFCAGAMAVGGALAYLVVKTLLINATOLLKLLAKLAEIVAAIAAD 180
QY 181 IISDVADIIGKTLGEVWEFTITNALNGELKELWDKLTGCVTGLFSRGWSNLESEFFAGVPGLT 240
Db 181 IISDVADIIGKTLGEVWEFTITNALNGELKELWDKLTGCVTGLFSRGWSNLESEFFAGVPGLT 240
QY 241 GATSGLSQVTLGFAAGLSASSGLAHADSLASSASLPALAGIGGSGFGLPSLAQVHAA 300
Db 241 GATSGLSQVTLGFAAGLSASSGLAHADSLASSASLPALAGIGGSGFGLPSLAQVHAA 300
QY 301 STRQALRPRADGPVGAARQVGGOSQVLSAQSOGMGVGVGGMHPSSGASKGTTTKY 360
Db 301 STRQALRPRADGPVGAARQVGGOSQVLSAQSOGMGVGVGGMHPSSGASKGTTTKY 360
QY 361 SEGAAAGTDAERAPVEADAGGQKVLVRNVV 392
Db 361 SEGAAAGTDAERAPVEADAGGQKVLVRNVV 392

RESULT 8
AAU01905
ID AAU01905 standard; Protein; 543 AA.
XX
AC AAU01905;
XX
XX
DT 29-AUG-2001 (first entry)
XX
XX
DE M. tuberculosis antigen HTCC#1 fusion protein #4.
XX
XX
KW TbrA12-HTCC#1; antigen; vaccine; tuberculosis;
KW AIDS; acquired immunodeficiency disease; His Tag.
XX
OS Mycobacterium tuberculosis.
OS Synthetic.
XX
XX
FH Key 3..8 Location/Qualifiers
FT Binding-site /label= "Histidine tag
FT /note= "Nickel chelating region used to aid
FT purification of the protein"
FT Misc-difference 541
FT /label= OTHER
FT /note= "in frame STOP codon"
XX
XX
PN WO200124820-A1.
XX
PD 12-APR-2001.
XX
XX
PF 10-OCT-2000; 2000WO-US28095.
XX
XX
PR 07-OCT-1999; 99US-0158338.
XX
XX
PR 07-OCT-1999; 99US-0158425.
XX
XX
PA (CORI-) CORIXA CORP.
XX
XX
PI Skeiky Y, Reed S, Houghton RL, McNeill PD, Dillon DC, Lodes ML;
XX
XX
DR WPI; 2001-290576/30.
XX
XX
PT Vaccinating against Mycobacteria infections in mammals using fusion
PT proteins comprising combinations of heterologous antigens
XX
XX
PS Claim 17; Fig 8; 168pp; English.
XX
XX
CC The sequence represents Mycobacterium tuberculosis fusion protein,
CC TbrA12-HTCC#1 and includes a His tag at the N-terminus to aid
CC purification. Compositions comprising at least 2 heterologous
CC antigens, as a fusion protein, and vectors expressing the fusion

CC proteins are used as vaccines to prophylactically immunise mammals
 CC (especially humans) against infection by Mycobacteria. The compositions
 CC contain at least 2 heterologous antigens that increase the serological
 CC sensitivity of individuals infected with tuberculosis, a disease
 CC frequently affecting patients with acquired immunodeficiency disease,
 CC AIDS.
 XX

XX Sequence 543 AA;

Query Match 99.7%; Score 1963; DB 22; Length 543;
 Best Local Similarity 99.7%; Pred. No. 1e-162;
 Matches 391; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSRAFIIDPTISAIDGLYDLIGIPNOGGILYSSLEYFEKALEELAAAPPGDGLGSA 60
 DB 149 MSRAFIIDPTISAIDGLYDLIGIPNOGGILYSSLEYFEKALEELAAAPPGDGLGSA 208
 QY 61 DKYAGKRNHNVPFQELADLDRLQLSLIHDOANAVQTRDILEGAKKGLFVRPVAVDLT 120
 DB 209 DKYAGKRNHNVPFQELADLDRLQLSLIHDOANAVQTRDILEGAKKGLFVRPVAVDLT 268
 QY 121 YIPVVGHALSAAFOAPFCAGAMVGGALAYLVVKTLLINATQLLKLAELVAAIA 180
 DB 269 YIPVVGHALSAAFOAPFCAGAMVGGALAYLVVKTLLINATQLLKLAELVAAIA 328
 QY 181 IISDVADIIGKTLGEVWEFITNALNGIKELWDKLTGWVTGLFSRGWSNLESPFAGVGLT 240
 DB 329 IISDVADIIGKTLGEVWEFITNALNGIKELWDKLTGWVTGLFSRGWSNLESPFAGVGLT 388
 QY 241 GATSGLSQVTLFGAAGLSASSGLAHADSLASSASLPALAGIGGSGFGGLPSLAQVHAA 300
 DB 389 GATSGLSQVTLFGAAGLSASSGLAHADSLASSASLPALAGIGGSGFGGLPSLAQVHAA 448
 QY 301 STRQALRPRADGPVGAAGVGGQSQQLVSAQSGQMGPGVGMGHPSPSSGASGKTTTKY 360
 DB 449 STRQALRPRADGPVGAAGVGGQSQQLVSAQSGQMGPGVGMGHPSPSSGASGKTTTKY 508
 QY 361 SEGAAAGTDEAPRVEADAGGQKVLVRNV 392
 DB 509 SEGAAAGTDEAPRVEADAGGQKVLVRNV 540

RESULT 9
 AAU01886
 ID AAU01886 standard; Protein; 408 AA.
 AC AAU01886;
 XX
 XX 29-AUG-2001 (first entry)
 XX
 XX M. tuberculosis antigen HTCC#1(TM-1).
 XX HTCC#1(TM-1); antigen; vaccine; tuberculosis; AIDS;
 XX acquired immunodeficiency disease; His Tag.
 XX
 XX Mycobacterium tuberculosis.
 XX
 XX Key Location/Qualifiers
 XX Binding-site 1..8
 XX /label= Histidine tag
 XX /note= "Nickel chelating region used to aid
 XX purification of the protein"
 XX Misc-difference 169..180
 XX /note= "These residues are deleted to form the
 XX HTCC#1(TM-1) protein sequence"
 XX Misc-difference 405
 XX /label= OTHER
 XX /note= "In frame STOP codon"
 XX
 XX WO200124820-A1.
 XX
 XX 12-APR-2001.
 XX

PF 10-OCT-2000; 2000WO-US28095.
 XX
 XX 07-OCT-1999; 99US-0158338.
 PR 07-OCT-1999; 99US-0158425.
 XX
 XX (CORI-) CORIXA CORP.
 XX
 XX Skeiky Y, Reed S, Houghton RL, McNeill PD, Dillon DC, Lodes ML;
 XX
 XX WPI; 2001-290576/30.
 DR N-PSDB; AAS03777.
 XX
 XX Vaccinating against Mycobacteria infections in mammals using fusion
 PT proteins comprising combinations of heterologous antigens -
 PT
 XX Example 2; Fig 9a; 168pp; English.
 XX
 XX The sequence is Mycobacterium tuberculosis HTCC#1(TM-1)
 CC representing HTCC#1 (an M. tuberculosis antigen) lacking the first
 CC transmembrane domain and includes a His tag at the N-terminus to aid
 CC purification. Compositions comprising at least 2 heterologous antigens,
 CC as a fusion protein, and vectors expressing the fusion proteins are used
 CC as vaccines to prophylactically immunise mammals (especially humans)
 CC against infection by Mycobacteria. The compositions contain at least 2
 CC heterologous antigens that increase the serological sensitivity of
 CC individuals infected with tuberculosis, a disease frequently affecting
 CC patients with acquired immunodeficiency disease, AIDS.
 XX
 XX Sequence 408 AA;
 SQ
 Query Match 98.9%; Score 1947; DB 22; Length 408;
 Best Local Similarity 99.0%; Pred. No. 1.7e-161;
 Matches 392; Conservative 0; Mismatches 0; Indels 4; Gaps 2;
 QY 1 MSRAFIIDPTISAIDGLYDLIGIPNOGGILYSSLEYFEKALEELAAAPPGDGLGSA 60
 DB 9 MSRAFIIDPTISAIDGLYDLIGIPNOGGILYSSLEYFEKALEELAAAPPGDGLGSA 68
 QY 61 DKYAGKRNHNVPFQELADLDRLQLSLIHDOANAVQTRDILEGAKKGLFVRPVAVDLT 120
 DB 69 DKYAGKRNHNVPFQELADLDRLQLSLIHDOANAVQTRDILEGAKKGLFVRPVAVDLT 128
 QY 121 YIPVVGHALSAAFOAPFCAGAMVGGALAYLVVKTLLINAKLTQLLKLAELVAA 176
 DB 129 YIPVVGHALSAAFOAPFCAGAMVGGALAYLVVKTLLINAKLTQLLKLAELVAA 188
 QY 177 AIADIISDVADIIGKTLGEVWEFITNALNGIKELWDKLTGWVTGLFSRGWSNLESPFAGV 236
 DB 189 AIADIISDVADIIGKTLGEVWEFITNALNGIKELWDKLTGWVTGLFSRGWSNLESPFAGV 248
 QY 237 PGLTGATSGLSQVTLFGAAGLSASSGLAHADSLASSASLPALAGIGGSGFGGLPSLAQ 296
 DB 249 PGLTGATSGLSQVTLFGAAGLSASSGLAHADSLASSASLPALAGIGGSGFGGLPSLAQ 308
 QY 297 VHAASRQALRPRADGPVGAAGVGGQSQQLVSAQSGQMGPGVGMGHPSPSSGASGKTT 356
 DB 309 VHAASRQALRPRADGPVGAAGVGGQSQQLVSAQSGQMGPGVGMGHPSPSSGASGKTT 368
 QY 357 TKKYSAGAAAGTDEAPRVEADAGGQKVLVRNV 392
 DB 369 TKKYSAGAAAGTDEAPRVEADAGGQKVLVRNV 404
 RESULT 10
 AAU01887
 ID AAU01887 standard; Protein; 408 AA.
 XX AAU01887;
 AC AAU01887;
 XX
 XX 29-AUG-2001 (first entry)
 XX
 XX M. tuberculosis antigen HTCC#1(TM-2).
 XX


```

XX DE M. tuberculosis antigen HTCC#1(1-223).
XX AC HTCC#1(1-223); antigen; vaccine; tuberculosis; AIDS;
XX AC acquired immunodeficiency disease; His Tag.
XX DT
XX OS Mycobacterium tuberculosis.
XX FH Key Location/Qualifiers
XX FT Binding-site 2..7 /label= Histidine tag
XX FT /note= "Nickel chelating region used to aid
XX FT purification of the protein"
XX FT Misc-difference 240 /label= OTHER
XX FT /note= "In frame STOP codon"
XX WO200124820-A1.
XX PD 12-APR-2001.
XX PF 10-OCT-2000; 2000WO-US28095.
XX PR 07-OCT-1999; 99US-0158338.
XX PR 07-OCT-1999; 99US-0158425.
XX PA (CORI-) CORIXA CORP.
XX PI Skeiky Y, Reed S, Houghton RL, McNeill PD, Dillon DC, Lodes ML;
XX DR WPI; 2001-290576/30.
XX PS Example 2; Fig 7a; 168pp; English.
XX CC The sequence is Mycobacterium tuberculosis HTCC#1(1-223),
XX CC representing the first transmembrane domain of HTCC#1 (an M.
XX CC tuberculosis antigen) and includes a His tag at the N-terminus to aid
XX CC purification. Compositions comprising at least 2 heterologous antigens,
XX CC as a fusion protein, and vectors expressing the fusion proteins are used
XX CC as vaccines to prophylactically immunise mammals (especially humans)
XX CC against infection by Mycobacterium. The compositions contain at least 2
XX CC heterologous antigens that increase the serological sensitivity of
XX CC individuals infected with tuberculosis, a disease frequently affecting
XX CC patients with acquired immunodeficiency disease, AIDS.
XX SQ Sequence 242 AA;
XX Query Match 59.6%; Score 1174; DB 22; Length 242;
XX Best Local Similarity 100.0%; Pred. No. 2.7e-94;
XX Matches 232; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSRAFIIDPTISAIDGLYLLGIGIPNOGGLYSSLEYFEKALEELAAFPDGHLSAA 60
Db 8 MSRAFIIDPTISAIDGLYLLGIGIPNOGGLYSSLEYFEKALEELAAFPDGHLSAA 67
QY 61 DKYAGKRNHNHVFQELADLDRLISLIHQANAVOTTRDILEGAKKGLFVRPVAVDLT 120
Db 68 DKYAGKRNHNHVFQELADLDRLISLIHQANAVOTTRDILEGAKKGLFVRPVAVDLT 127
QY 121 YIPVWGHLSAAFPQAFPCAGAMAVGGLAYLVVKTLLINATOLLKLLAKLAEVAAATAD 180
Db 128 YIPVWGHLSAAFPQAFPCAGAMAVGGLAYLVVKTLLINATOLLKLLAKLAEVAAATAD 187
QY 181 IISDVADI IKGTLGEVWEFITNALNGKELWDKLTGWTGLFSRGWSNLESP 232
Db 188 IISDVADI IKGTLGEVWEFITNALNGKELWDKLTGWTGLFSRGWSNLESP 239
XX RESULT 14

```

```

AAU01884
XX ID AAU01884 standard; Protein; 219 AA.
XX AC AAU01884;
XX DT 29-AUG-2001 (first entry)
XX DE M. tuberculosis antigen HTCC#1(184-392).
XX FH Key Location/Qualifiers
XX FT Binding-site 2..7 /label= Histidine tag
XX FT /note= "Nickel chelating region used to aid
XX FT purification of the protein"
XX FT Misc-difference 216 /label= OTHER
XX FT /note= "In frame STOP codon"
XX WO200124820-A1.
XX PD 12-APR-2001.
XX PF 10-OCT-2000; 2000WO-US28095.
XX PR 07-OCT-1999; 99US-0158338.
XX PR 07-OCT-1999; 99US-0158425.
XX PA (CORI-) CORIXA CORP.
XX PI Skeiky Y, Reed S, Houghton RL, McNeill PD, Dillon DC, Lodes ML;
XX DR WPI; 2001-290576/30.
XX PS Vaccinating against Mycobacteria infections in mammals using fusion
XX FT proteins comprising combinations of heterologous antigens -
XX CC Example 2; Fig 7b; 168pp; English.
XX CC The sequence is Mycobacterium tuberculosis HTCC#1(184-392)
XX CC representing the two C-terminal transmembrane domains of HTCC#1 (a M.
XX CC tuberculosis antigen) and includes a His tag at the N-terminus to aid
XX CC purification. Compositions comprising at least 2 heterologous antigens,
XX CC as a fusion protein, and vectors expressing the fusion proteins are used
XX CC as vaccines to prophylactically immunise mammals (especially humans)
XX CC against infection by Mycobacterium. The compositions contain at least 2
XX CC heterologous antigens that increase the serological sensitivity of
XX CC individuals infected with tuberculosis, a disease frequently affecting
XX CC patients with acquired immunodeficiency disease, AIDS.
XX SQ Sequence 219 AA;
XX Query Match 53.9%; Score 1061; DB 22; Length 219;
XX Best Local Similarity 100.0%; Pred. No. 1.7e-84;
XX Matches 209; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 184 DVADI IKGTLGEVWEFITNALNGKELWDKLTGWTGLFSRGWSNLESPFAGVPGTGTAT 243
Db 8 DVADI IKGTLGEVWEFITNALNGKELWDKLTGWTGLFSRGWSNLESPFAGVPGTGTAT 67
QY 244 SGLSQVTGLFGAAGLSSGLAHADSLASSASLPALAGICGGSGFGLPSLAOVHAASTR 303
Db 68 SGLSQVTGLFGAAGLSSGLAHADSLASSASLPALAGICGGSGFGLPSLAOVHAASTR 127
QY 304 QALPRADGPVGAAGAAEQVGSQSLVSAQSGQGMGGPVGMGMGHPSSGASKGTTTKYSEG 363
Db 128 QALPRADGPVGAAGAAEQVGSQSLVSAQSGQGMGGPVGMGMGHPSSGASKGTTTKYSEG 187
QY 364 AAAGTEDAERAPVEADAGGGQKVLVRNV 392

```

```

Db      69  SGLSQVTGLFGAAGLSASSGLAHADSLASSLFPALAGIGGGSGFGLPSLAQVHAASR 128
QY      304  QALPRADGPVGAARQVGGOSQLVSAQSGQMGPPVGMGMHPPSSGASKGTTTKYSEG 363
Db      129  QALPRADGPVGAARQVGGOSQLVSAQSGQMGPPVGMGMHPPSSGASKGTTTKYSEG 188
QY      364  AAAGTEDAERAPVEADAGGQKVLNVVV 392
Db      189  AAAGTEDAERAPVEADAGGQKVLNVVV 217

Search completed: November 5, 2003, 16:59:50
Job time : 53 secs

```

Db	188	AAAGTDAERAPVEADAGGQKVLVRNV	216
RESULT 15			
AAU01902			
ID	AAU01902	standard; Protein; 744 AA.	
XX	AAU01902;		
XX	29-AUG-2001	(first entry)	
DE	M. tuberculosis antigen HTCC#1	fusion protein #1.	
XX	HTCC#1(184-392)-TbH9-HTCC#1(1-129);	antigen; vaccine; tuberculosis;	
KW	AIDS; acquired immunodeficiency disease;	His Tag.	
XX	Mycobacterium tuberculosis.		
OS	Synthetic.		
XX	Key	Location/Qualifiers	
PH	Binding-site	3..8	
FT	/label= Histidine tag		
FT	/note= "Nickel chelating region used to aid		
FT	purification of the protein"		
FT	Misc-difference	742	
FT	/label= OTHER		
FT	/note= "In frame STOP codon"		
XX	WO200124820-A1.		
PN	12-APR-2001.		
XX	10-OCT-2000; 2000WO-US28095.		
PD	07-OCT-1999; 99US-0158338.		
XX	07-OCT-1999; 99US-0158425.		
PR	(CORI-) CORIXA CORP.		
ER	Skeiky Y, Reed S, Houghton RL, McNeill PD, Dillon DC, Lodes ML;		
XX	WPI; 2001-290576/30.		
DR	N-PSDB; AAS03790.		
XX	Vaccinating against Mycobacteria infections in mammals using fusion		
PT	proteins comprising combinations of heterologous antigens -		
XX	Claim 14; Fig 10; 168pp; English.		
PS	The sequence represents Mycobacterium tuberculosis fusion protein,		
XX	HTCC#1(184-392)-TbH9-HTCC#1(1-129) and includes a His tag at the		
CC	N-terminus to aid purification. Compositions comprising at least 2		
CC	heterologous antigens, as a fusion protein, and vectors expressing the		
CC	fusion proteins are used as vaccines to prophylactically immunise mammals		
CC	(especially humans) against infection by Mycobacteria. The compositions		
CC	contain at least 2 heterologous antigens that increase the serological		
CC	sensitivity of individuals infected with tuberculosis, a disease		
CC	frequently affecting patients with acquired immunodeficiency disease,		
CC	AIDS.		
XX	Sequence	744 AA;	
SQ	Query Match	53.6%; Score 1055; DB 22; Length 744;	
	Best Local Similarity	99.5%; Pred. No. 3e-83;	
	Matches 208; Conservative	0; Mismatches 1; Indels 0; Gaps 0;	
QY	184	DVADIIKGLGVEWFFINALNGKLKELDKLTGVTGLFSGRGSNLESFFAGVPGLTGAT	243
Db	9	DVADIIKGLGVEWFFINALNGKLKELDKLTGVTGLFSGRGSNLESFFAGVPGLTGAT	68
QY	244	SLGSQVTGLFGAAGLSASSGLAHADSLASSASLPALACIGGGSGFGLPELAQVHAASR	303

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 5, 2003, 17:00:27 ; Search time 21 Seconds
(without alignments)
789.803 Million cell updates/sec

Title: US-09-688-672A-14
Perfect score: 1969
Sequence: 1 MSRAFIIDPTISAIDGLYDL.....RAPVEADAGGOKVLVRNV 392

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2.6/prodata/2/iaa/5A_COMB.pep.*
- 2: /cgn2.6/prodata/2/iaa/5B_COMB.pep.*
- 3: /cgn2.6/prodata/2/iaa/6A_COMB.pep.*
- 4: /cgn2.6/prodata/2/iaa/8B_COMB.pep.*
- 5: /cgn2.6/prodata/2/iaa/PCUS_COMB.pep.*
- 6: /cgn2.6/prodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	1969	100.0	392	US-09-073-009-138	Sequence 138, Appl
2	148.5	7.5	269	US-09-134-001C-3461	Sequence 3461, Ap
3	135	6.9	318	US-09-060-756-727	Sequence 727, App
4	135	6.9	318	US-09-670-314-727	Sequence 727, App
5	133.5	6.8	731	US-09-340-736E-1	Sequence 1, Appli
6	133	6.8	1077	US-07-972-032-82	Sequence 82, Appl
7	133	6.8	1077	US-08-642-255-95	Sequence 95, Appl
8	129.5	6.6	731	US-08-911-364-1	Sequence 1, Appli
9	129.5	6.6	733	US-08-464-700-2	Sequence 2, Appli
10	129.5	6.6	1271	US-08-095-734-2	Sequence 2, Appli
11	129.5	6.6	1271	US-08-444-623-2	Sequence 2, Appli
12	129.5	6.6	1271	US-08-471-869-2	Sequence 2, Appli
13	129.5	6.6	1271	US-09-342-563-2	Sequence 2, Appli
14	129.5	6.6	1271	PCN-US94-08267-2	Sequence 2, Appli
15	128.5	6.5	832	US-08-209-747-2	Sequence 2, Appli
16	128.5	6.5	832	US-08-458-298-2	Sequence 2, Appli
17	126.5	6.4	334	US-09-060-756-728	Sequence 728, App
18	126.5	6.4	334	US-09-670-314-728	Sequence 728, App
19	126	6.4	1958	US-07-945-283-2	Sequence 2, Appli
20	125	6.3	738	US-08-864-038A-3	Sequence 3, Appli
21	125	6.3	2088	US-09-548-372D-13	Sequence 13, Appl
22	125	6.3	2088	US-09-548-367D-13	Sequence 13, Appl
23	125	6.3	2088	US-09-551-853D-13	Sequence 13, Appl
24	124.5	6.3	1415	US-09-252-991A-26438	Sequence 26438, A
25	124	6.3	792	US-08-678-039A-40	Sequence 40, Appl
26	124	6.3	1056	US-08-212-237-6	Sequence 6, Appli
27	124	6.3	1056	US-08-806-029-29	Sequence 29, Appl

28	124	6.3	1056	5	PCT-US95-02772-6	Sequence 6, Appli
29	124	6.3	1136	4	US-08-806-029-9	Sequence 9, Appli
30	123.5	6.3	651	3	US-08-556-978B-19	Sequence 19, Appli
31	123.5	6.3	651	3	US-09-247-806-1	Sequence 1, Appli
32	123.5	6.3	718	1	US-08-425-069-2	Sequence 2, Appli
33	123.5	6.3	718	2	US-08-317-844B-2	Sequence 2, Appli
34	123.5	6.3	747	3	US-09-034-177-3	Sequence 3, Appli
35	122	6.2	159	1	US-07-609-716-104	Sequence 104, App
36	122	6.2	159	3	US-08-475-411A-104	Sequence 104, App
37	122	6.2	159	3	US-08-478-029A-104	Sequence 104, App
38	121.5	6.2	1694	1	US-08-494-168-2	Sequence 2, Appli
39	121	6.1	785	4	US-03-252-991A-32952	Sequence 32952, A
40	120	6.1	604	3	US-08-556-978B-63	Sequence 63, Appli
41	120	6.1	766	1	US-08-175-155-53	Sequence 53, Appli
42	120	6.1	766	1	US-08-477-509B-88	Sequence 88, Appli
43	120	6.1	766	2	US-08-707-237A-61	Sequence 61, Appli
44	120	6.1	766	3	US-08-482-085B-88	Sequence 88, Appli
45	120	6.1	766	4	US-09-444-791A-88	Sequence 88, Appli

ALIGNMENTS

RESULT 1
US-09-073-009-138
; Sequence 138 Application US/09073009
; Patent No. 6555853
; GENERAL INFORMATION:
; APPLICANT: Alderson, Mark
; APPLICANT: Dillon, Davin C.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Campos-Neto, Antonio
; TITLE OF INVENTION: COMPOUNDS AND DIAGNOSIS OF
; NUMBER OF SEQUENCES: 144
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Ave.
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/09/073.009
; APPLICATION NUMBER: 05-MAY-1998
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Mak, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/POCKET NUMBER: 210121.441C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-622-4900
; TELEFAX: 206-682-6031
; INFORMATION FOR SEQ ID NO: 138:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 392 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-073-009-138

Query Match 100.0%; Score 1969; DB 4; Length 392;
Best Local Similarity 100.0%; Pred. No. 1.3e-175;
Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSRAFIIDPTISAIDGLYDLIGIPNQGILYSSLEYFEKALEELAAAPFGDWLGSA 60
DB 1 MSRAFIIDPTISAIDGLYDLIGIPNQGILYSSLEYFEKALEELAAAPFGDWLGSA 60

TUBERCULOSIS AND ME

QY 61 DKYAGKRNHNFFQBLADLDROLISLHDOANVQTRDILEGAKKGLFVRPVAVDLT 120
 Db 61 DKYAGKRNHNFFQBLADLDROLISLHDOANVQTRDILEGAKKGLFVRPVAVDLT 120
 QY 121 YIPVVGHLSAAFAOPFCAGAMVVGALAYLVKTLINATOLLKLLAKLAELVAAAIAAD 180
 Db 121 YIPVVGHLSAAFAOPFCAGAMVVGALAYLVKTLINATOLLKLLAKLAELVAAAIAAD 180
 QY 181 IISDVADIITKTLGEVWEFTITNALNGELKELDKLTGWVTGLFSRGSNLSFFAGVPGLT 240
 Db 181 IISDVADIITKTLGEVWEFTITNALNGELKELDKLTGWVTGLFSRGSNLSFFAGVPGLT 240
 QY 241 GATSGLSQVTLGFGAAGLSASSGLAHADSLASSASLPALAGIGGSGFGGLPSLAQVHAA 300
 Db 241 GATSGLSQVTLGFGAAGLSASSGLAHADSLASSASLPALAGIGGSGFGGLPSLAQVHAA 300
 QY 301 STROALRPRADPGVGAARQVGGQSLVSAQSGQMGPGVGMGMHPSSGASKGTTTKY 360
 Db 301 STROALRPRADPGVGAARQVGGQSLVSAQSGQMGPGVGMGMHPSSGASKGTTTKY 360
 QY 361 SEGAAAGTDEADRAPVEADAGGQKVLVRNVV 392
 Db 361 SEGAAAGTDEADRAPVEADAGGQKVLVRNVV 392

RESULT 2
 US-09-134-001C-3461
 ; Sequence 3461, Application US/09134001C
 ; Patent No. 6330370
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn Doucette-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: GTC-007
 ; CURRENT APPLICATION NUMBER: US/09/134,001C
 ; CURRENT FILING DATE: 1998-08-13
 ; PRIOR APPLICATION NUMBER: US 60/064,964
 ; PRIOR FILING DATE: 1997-11-08
 ; PRIOR APPLICATION NUMBER: US 60/055,779
 ; PRIOR FILING DATE: 1997-08-14
 ; NUMBER OF SEQ ID NOS: 5674
 ; SEQ ID NO 3461
 ; LENGTH: 269
 ; TYPE: PRT
 ; ORGANISM: Staphylococcus epidermidis
 US-09-134-001C-3461

Query Match 7.5%; Score 148.5; DB 4; Length 269;
 Best Local Similarity 34.0%; Pred. No. 6.8e-06;
 Matches 48; Conservative 18; Mismatches 66; Indels 9; Gaps 4;
 QY 234 AGVPGLTG--ATSGLSQVTLGFGAAGLSASSGLAHADSLASSASLPALAGIGGSGF--- 288
 Db 20 AGVPGSTGLPGSTGLPDSAGLPGSAGVPGSAGLPGSAGVPGSAGLPGSAGLPGS 79
 QY 289 GGLPSLAQV--HAASRQALRPRADPGVGAARQVGGQSLVSAQSGQMGPGVGMGMH 346
 Db 80 AGVPGSAGVPGSAGLPGSAGLPGSAGLPGSAG--VPGSAGLPGSAGVPGSAGLPGSAGLP 137
 QY 347 PSSGASKGTTTKYSEGAAG 367
 Db 138 GSAGLPGSAFVPGYSTISSTG 158

RESULT 3
 US-09-080-756-727
 ; Sequence 727, Application US/09060756
 ; Patent No. 6183957
 ; GENERAL INFORMATION:
 ; APPLICANT: Cole, Stewart
 ; APPLICANT: Buchrieser-Brosch, Roland
 ; APPLICANT: Gordon, Stephen

; APPLICANT: Billault, Alain
 ; TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM
 ; TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
 ; TITLE OF INVENTION: LIBRARY APPLICATION TO THE DETECTION OF MYCOBACTERIA
 ; FILE REFERENCE: 3495-0169
 ; CURRENT APPLICATION NUMBER: US/09/060,756
 ; CURRENT FILING DATE: 1998-04-16
 ; NUMBER OF SEQ ID NOS: 743
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 727
 ; LENGTH: 318
 ; TYPE: PRT
 ; ORGANISM: Mycobacterium sp.
 US-09-060-756-727

Query Match 6.9%; Score 135; DB 3; Length 318;
 Best Local Similarity 27.5%; Pred. No. 0.00016;
 Matches 58; Conservative 11; Mismatches 104; Indels 38; Gaps 7;
 QY 191 GTLGEVWEFTITNALNGELKELDKLTGWVTGLFSRGSNLSFFAGVPGLTGATSGLSQVT 250
 Db 14 GTPCAGVSGATGAPG-----GWLIGDGGAGGGAAG--SGAPGAGGAAGLWGTG 61
 QY 251 GLFGAAGLSASSG-----LAHADSLASSASLPALAGIGGSGFGGLPSLAQVH 298
 Db 62 GAGGAGSSAGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 121
 QY 299 AASTROALRPRADPGVGAARQVGGQSLVSAQSGQ-----MGPGVGMGMHPSSGAS 352
 Db 122 GAGGTGLV--GGDGGAGGAGGTGGLLAGLIGAGGCHGTGGLSTNGDGGVGG---AGCN 175
 QY 353 KCTTKKYSEGAAGTDEADRAPVEADAGG 383
 Db 176 AGMLAPGGAGGAGG--DGENLDTGGDGGAG 204

RESULT 4
 US-09-670-314-727
 ; Sequence 727, Application US/09670314
 ; Patent No. 6492506
 ; GENERAL INFORMATION:
 ; APPLICANT: Cole, Stewart
 ; APPLICANT: Buchrieser-Brosch, Roland
 ; APPLICANT: Billault, Alain
 ; TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM
 ; TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
 ; TITLE OF INVENTION: LIBRARY APPLICATION TO THE DETECTION OF MYCOBACTERIA
 ; FILE REFERENCE: 3495-0169
 ; CURRENT APPLICATION NUMBER: US/09/670,314
 ; CURRENT FILING DATE: 2001-01-12
 ; PRIOR APPLICATION NUMBER: 09/060,756
 ; PRIOR FILING DATE: 1998-04-16
 ; NUMBER OF SEQ ID NOS: 743
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 727
 ; LENGTH: 318
 ; TYPE: PRT
 ; ORGANISM: Mycobacterium sp.
 US-09-670-314-727

Query Match 6.9%; Score 135; DB 4; Length 318;
 Best Local Similarity 27.5%; Pred. No. 0.00016;
 Matches 58; Conservative 11; Mismatches 104; Indels 38; Gaps 7;
 QY 191 GTLGEVWEFTITNALNGELKELDKLTGWVTGLFSRGSNLSFFAGVPGLTGATSGLSQVT 250
 Db 14 GTPCAGVSGATGAPG-----GWLIGDGGAGGGAAG--SGAPGAGGAAGLWGTG 61
 QY 251 GLFGAAGLSASSG-----LAHADSLASSASLPALAGIGGSGFGGLPSLAQVH 298
 Db 62 GAGGAGSSAGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 121

QY 299 AASTRALPRADGPVGAARQVGGQSLVSAQSGG-----MGPVGMGMHPSSGAS 352
Db 122 GAGTGLV--GGDGGAGGAGTGGLLAGLIGAGGGHGTGGLSTNGDGGVGG-----AGGN 175
QY 353 KGTITTKYSEGAAGTEDAERAPVEADAGG 383
Db 176 AGWLAGPGGAGGAG--DGENLDTGGDGGAG 204

RESULT 5
US-09-340-736E-1
; Sequence 1, Application US/09340736E
; Patent No. 6489446
; GENERAL INFORMATION:
; APPLICANT: ROTHSTEIN, ASER
; APPLICANT: KEELEY, FRED
; APPLICANT: ROTHSTEIN, STEVEN
; TITLE OF INVENTION: SELF-ALIGNING PEPTIDES MODELED ON HUMAN ELASTIN
; TITLE OF INVENTION: AND OTHER FIBROUS PROTEINS
; FILE REFERENCE: 041082/0110
; CURRENT APPLICATION NUMBER: US/09/340,736E
; CURRENT FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: 08/911,364
; PRIOR FILING DATE: 1997-08-07
; PRIOR APPLICATION NUMBER: 60/023,552
; PRIOR FILING DATE: 1996-08-07
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patent Ver. 2.1
; SEQ ID NO 1
; LENGTH: 731
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-340-736E-1

Query Match 6.8%; Score 133.5; DB 4; Length 731;
Best Local Similarity 20.4%; Pred. No. 0.00074;
Matches 86; Conservative 35; Mismatches 136; Indels 165; Gaps 15;

QY 22 GIGIPN--CCGILYSSLEYFEKALELEAAFFPGDGLWSAADKYAGKKNHNVFFQELAD 79
Db 117 GVGLPGVPGVVL-----PGARFPGVLPGV-----143
QY 80 LDRQLISLHDQANAVQTRDILEGAKKLEFVRPV-----AVDLTYTPVVGHALSAAPQ 134
Db 144 -----PTGAGVKPKAPGVGGAGFAGIPGVPGFPQPGVPLGY-PIKAPKLPGGY 192
QY 135 APCAGAM-----AVVGA-LAYLVVKTLLNATOLLKLLAKLAELVAATADIIISDA 186
Db 193 LPTTKLPGYGVPGVGAAGKAGYPTGTGVGPQMAAATAAKAAGAGVLPVGV 252
QY 187 DIKGTLGEVWFITNALNGLKELMDKLTGWVTGLFSRGWSNLESFFAGVPLGTGATSL 246
Db 253 G-----AGVPGVPGA-----262
QY 247 SQVTGLFGAAGLSASSGLAHADSLASSASLPALAG-IGGSGFG-----GLPSL-- 294
Db 263 --IPGGTGGTGTPTAAAAAATAAKAAGAGLVPGGPGFPGVGVPGAGVPGVGV 320
QY 295 -----AQVHAASTRQALRPRADGPVGAARQVGG-----322
Db 321 PGAGIPVPGAGIPGAAGVGVVSPFAAATAAKAAGKAGARPGVGVGGIPTYGVGAGFP 380
QY 323 ----GQSOLVSAQSGQMGVPMG-----GMPSSGASGKTTTKYSEG--AAATEDAE 372
Db 381 FGVGVGGIPGVAGVPGVGVGGVPGVGVGIPGAGVGVGIPGAGVGVGTPAAAAAKAAA 440
QY 373 RA 374
Db 441 KA 442

RESULT 6
US-07-972-032-82

; Sequence 82, Application US/07972032
; Patent No. 5496712
; GENERAL INFORMATION:
; APPLICANT: Cappello, Joseph
; APPLICANT: Ferrari, Franco A.
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT COLLAGEN-LIKE
; TITLE OF INVENTION: PROTEIN POLYMERS
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bertram I. Rowland
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: CA 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/07/972,032
; APPLICATION NUMBER: 07/972,032
; FILING DATE: 19921105
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/791,960
; FILING DATE: 12-NOV-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Bertram I.
; REGISTRATION NUMBER: 20,015
; REFERENCE/DOCKET NUMBER: A-58556-1/BIR/PROP-08-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; INFORMATION FOR SEQ ID NO: 82:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1077 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-972-032-82

Query Match 6.8%; Score 133; DB 1; Length 1077;
Best Local Similarity 31.9%; Pred. No. 0.0015;
Matches 51; Conservative 10; Mismatches 85; Indels 14; Gaps 5;
QY 235 GVPLTGA-----TSGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGICGSGFG 289
Db 43 GAPLQAGAPGAPGSGAGLQAGAPGAPGSGAGLQAGAPGAPGSGAGLQAGAPGAPGSG 102
QY 290 GLPSLAQVHAASTRQALRPRADGPVGAARQVGGQSLVSAQSGQMGVPMGMHPSS 349
Db 103 GAPLQAGAPGAPGSG--APLQAGAPGAPGSG--GAPLQAGAPGAPGSGAGLQAG- 158
QY 350 GASKGTTTKYSEGA-----AAGTEDAERAPVEADAGGQ 384
Db 159 PGSGAPGLQAGAPGAPGSGAGLQAGAPGAPGSGAGLQAGAPGAPGSGAGLQ 198

RESULT 7
US-08-642-255-95
; Sequence 95, Application US/08642255
; Patent No. 5773249
; GENERAL INFORMATION:
; APPLICANT: Cappello, Joseph
; APPLICANT: Ferrari, Franco A.
; TITLE OF INVENTION: High Molecular Weight Collagen-Like
; TITLE OF INVENTION: Protein Polymers
; NUMBER OF SEQUENCES: 135
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400

Db 1234 GVRFCRSVGGPGLGACAGACCTAGPLG 1261

RESULT 11

US-08-444-623-2

Sequence 2, Application US/08444623

Patent No. 5866403

GENERAL INFORMATION:

APPLICANT: Aldovini, Anna

APPLICANT: Young, Richard A.

TITLE OF INVENTION: Homologously Recombinant Slow Growing

TITLE OF INVENTION: Mycobacteria and Uses Therefor

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.

STREET: Two Militia Drive

CITY: Lexington

STATE: MA

COUNTRY: USA

ZIP: 02173

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/444,623

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/095,734

FILING DATE: 22-JUL-1993

APPLICATION NUMBER: 07/711,334

FILING DATE: 06-JUN-1991

APPLICATION NUMBER: 07/367,894

FILING DATE: 19-JUN-1989

APPLICATION NUMBER: PCT/US90/03451

FILING DATE: 18-JUN-1990

APPLICATION NUMBER: PCT/US89/02962

FILING DATE: 07-JUL-1989

APPLICATION NUMBER: 07/361,944

FILING DATE: 05-JUN-1989

APPLICATION NUMBER: 07/223,089

FILING DATE: 22-JULY-1988

APPLICATION NUMBER: 07/216,390

FILING DATE: 07-JUL-1988

APPLICATION NUMBER: 07/163,546

FILING DATE: 03-MAR-1988

APPLICATION NUMBER: PCT/US88/00614

FILING DATE: 29-FEB-1988

APPLICATION NUMBER: 07/020,451

FILING DATE: 02-MAR-1987

ATTORNEY/AGENT INFORMATION:

NAME: Granahan, Patricia

REGISTRATION NUMBER: 32,227

REFERENCE/DOCKET NUMBER: WH93-11M2

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-861-6240

TELEFAX: 617-861-9540

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1271 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-444-623-2

Query Match 6.6%; Score 125.5; DB 2; Length 1271;

Best Local Similarity 21.4%; Pred. No. 0.0039;

Matches 96; Conservative 52; Mismatches 155; Indels 145; Gaps 19;

22 GIGIPNQGILYS-----SLEYFEKA--LEELAAAFPGDGNLGSAAADKYAGKRNHVN 72

QY

831 GAGV---GPM LISATQNGELSAQYAAEAASEVELLGVWASEGQGAALVAAYMPFLA 887

73 FF-----QELADLRQLISLIHQDQANAVQTRDLEGAKKGLFVRVPVAVDLTIPVVGHA 128

888 WLIQASADCVEMAAQQAQHAVIEAYTAVALMPTQVELAANQIKLAVLVA TNFFGINTIPIA 947

129 LSAAFQAPFCAGAMAVVGGALAVLVKTLINA--TOLLKLLAKLAELVAAAIAIDISD- 184

948 INEAIEVEMVRAATM--ATYSTVSRSAISAPHTSPPLILKSDLELLPDTCGDSDEG 1005

185 -----VADIKKG-TLGE-VWEFITNALNGL-----KELWDKUT 215

1006 HNEGSHGHGARMIDNFAELIRGVSAGRIVMDPVNGTLNGLDYDDYVYPGHAIW---- 1061

216 GWYT-GL--FSRGSNLSFAGVPG----- 238

1062 -WLARGLEFFDQGEQFGELLFTNPTGAFQFLLYVVVVVDLPTHTAQIATWLGQVPQLLSAA 1120

239 LTGATSGLSQVTGLFGAAGLSASSGLA-----HADSLASSASLPALAGIG----- 283

1121 LTGVIAHLGAITGLAGLSGLSAIPSAIPAIVAVPELTPVAAAPMLAVAGVGAAPGML 1180

284 -----GSGFGGLPSLAQVHAASRAQLEPRADGPGVGAABQVG- 322

1181 PASAPAPAAAAGATAAGTPFPATGFGGLPAL-----PGRWRPRNRVRLGTGVPGRQGR 1233

323 GQSQVLSAQSQGMG-----GPVG 341

1234 GVRFCRSVGGPGLGACAGACCTAGPLG 1261

RESULT 12

US-08-471-869-2

Sequence 2, Application US/08471869

Patent No. 6022745

GENERAL INFORMATION:

APPLICANT: Aldovini, Anna

APPLICANT: Young, Richard A.

TITLE OF INVENTION: Homologously Recombinant Slow Growing

TITLE OF INVENTION: Mycobacteria and Uses Therefor

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.

STREET: Two Militia Drive

CITY: Lexington

STATE: MA

COUNTRY: USA

ZIP: 02173

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/471,869

FILING DATE: 07-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/US94/08267

FILING DATE: 22-JUL-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/095,734

FILING DATE: 22-JUL-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/711,334

FILING DATE: 06-JUN-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/367,894

FILING DATE: 19-JUN-1989

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/US90/03451

FILING DATE: 18-JUN-1990


```

; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Homologously Recombinant Slow Growing
; TITLE OF INVENTION: Mycobacteria and Uses Therefor
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/08267
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/095,734
; FILING DATE: 22-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: WH93-11MA PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1271 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US94-08267-2

```

```

Query Match 6.6%; Score 129.5; DB 5; Length 1271;
Best Local Similarity 21.4%; Pred. No. 0.0039;
Matches 96; Conservative 52; Mismatches 155; Indels 145; Gaps 19;

QY 22 GIGPQGGILYS-----SLEYSEKA--LEELAAFPDGLGSAADKYAGKUNHVN 72
DB 831 GAGV---GPMILSATONGELSAQYEAASEVEELLGVVASEGQQAALVAAYMPFLA 887
QY 73 FF----QELADLRQLISLHDQANAVQTRDILEGAKKGLFVFPVAVDLTYIPVVGHA 128
DB 888 WLQASADCEVAAQOAHVIAEYATRAVELMPTQVELAANOIKLAVLVATNFFGINTPIA 947
QY 129 LSAAFQAPFCAGMAMVVGALAYLVVKTLNA---TQLLKLLAKLAEIVAAAIADISD- 184
DB 948 INEAEEVEMVRAATTM--ATYSTVSRSALSAMPHTSPPPLILKSDLELLPTGDSDDG 1005
QY 185 -----VADIKKG-TLGE-VWEFITNALNGL-----KELWDLKT 215
DB 1006 HNHGGHSHGHARMIDNFAELIRGVSAAGRIVWDFVNGTLNGLDLDVYVPGHAIW---- 1061
QY 216 GWVT-GL--FSRGWNLSEFFAGVPG----- 238
DB 1062 -WLARGLEFFQDGEFGELLFNPFCAPQLLYVVVVDLPHIAQIATWLQGYQLLSAA 1120
QY 239 LFGATSGLSQVTLFGAAGLSASSGLA-----HADSLASSASLPALAGIG----- 283
DB 1121 LTVIAHLGAIITGLAGLSGLSAIPSAIIPAVVPELTVPVAAAPPMLAVAGVPAVAPGML 1180
QY 284 -----GGSGFGGLPSLAQVHAASTRALPRADGPPVGAABQVG- 322
DB 1181 PASAPAPAAAAGATAAGPTTPATGFGGLPAL-----PGRWRPNNRVRLGTVPGRQR 1233
QY 323 GQSQLVSAQGSQGMG-----GPVG 341

```

```

DB 1234 GVRFRCSRVGGPGLGACAGACCTAGPLG 1261
RESULT 15
US-08-209-747-2
; Sequence 2, Application US/08209747
; Patent No. 5733771
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph V.
; APPLICANT: Colgin, Mark
; TITLE OF INVENTION: CDNAs Encoding Minor Ampullate Spider
; TITLE OF INVENTION: Silk Proteins
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch
; STREET: P.O. Box 747
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22040-3487
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/209,747
; FILING DATE: 14-MAR-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M.
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1447-104P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-205-8000
; TELEFAX: 703-205-8050
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 832 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: N. clavipes
; TISSUE TYPE: minor ampullate gland
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1...309
; US-08-209-747-2

Query Match 6.5%; Score 128.5; DB 1; Length 832;
Best Local Similarity 24.6%; Pred. No. 0.0026;
Matches 70; Conservative 28; Mismatches 129; Indels 57; Gaps 11;

QY 121 YIPVGHASAPQAPFCAGMAMVVGALAYLVVKTLINATQLLKLLAKLAEIVAAAIAD 180
DB 11 YVTLGNALSDA-----SAYANALUSSAIGNVLANS-----GSISSSTASSAA- 52
QY 181 IISDVADIINGTL---GEVWEFITNALNGLKELWDLKTGWTG---LFSRGWNLSEFFA 234
DB 53 --SSAASSVTTLTSTGYPAVFVAPSAGS-----CYGAGAGAVAAAGAGAGGYGR 101
QY 235 GVPGLTCAISGISOVTLFGAAGLSASSGLAHADSLASSASLPALAGIG---GGSGFGGL 291
DB 102 GAGGY--GGGGYVAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 160
QY 292 PSIAQVHAASTRALPRADGPPVG-----AAAEQVGGQSOLVSAQGSQGMGPPV 340
DB 161 GGYGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 214

```


This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 5, 2003, 17:02:43 / Search time 44 Seconds
(without alignments)

1530.130 Million cell updates/sec

Title: US-09-688-672A-14

Perfect score: 1969

Sequence: 1 MSRAFDITSAIDGLYDL.....RAPVEADAGGQKVLVRNVV 392

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 644079 seqs, 171749292 residues

Total number of hits satisfying chosen parameters: 644079

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1969	100.0	392	9	US-09-073-009-138
2	1969	100.0	392	9	US-09-793-306-138
3	1969	100.0	392	12	US-10-098-732A-29
4	1969	100.0	392	16	US-10-080-170-610
5	1258	63.9	394	16	US-10-080-170-53
6	140.5	7.1	256	11	US-09-820-843A-18
7	135	6.9	318	12	US-10-259-678-727
8	133.5	6.8	731	12	US-09-964-662-1
9	131.5	6.7	446	15	US-10-156-761-13993
10	130	6.6	1079	11	US-09-820-843A-20
11	129.5	6.6	730	11	US-09-961-403-8
12	129	6.6	1149	12	US-10-097-111-283
13	127	6.4	585	10	US-09-738-626-6032
14	126.5	6.4	334	12	US-10-259-678-728
15	126	6.4	749	15	US-10-156-761-11729

16 123.5 6.3 651 9 US-09-861-597-1
17 123.5 6.3 651 12 US-10-414-760-1
18 121.5 6.2 1488 10 US-09-738-826-5495
19 120.5 6.1 1321 15 US-10-184-644-371
20 120.5 6.1 1321 15 US-10-184-644-371
21 120 6.1 766 15 US-10-096-986-88
22 120 6.1 979 15 US-10-096-986-88
23 120 6.1 1894 12 US-10-159-722-97
24 120 6.1 1894 12 US-10-187-749-97
25 120 6.1 1894 12 US-10-134-457-97
26 120 6.1 1894 12 US-10-184-642-97
27 120 6.1 1894 12 US-10-196-747-97
28 120 6.1 1894 12 US-10-173-689-97
29 120 6.1 1894 12 US-10-173-690-97
30 120 6.1 1894 12 US-10-173-691-97
31 120 6.1 1894 12 US-10-173-692-97
32 120 6.1 1894 12 US-10-173-694-97
33 120 6.1 1894 12 US-10-173-698-97
34 120 6.1 1894 12 US-10-173-699-97
35 120 6.1 1894 12 US-10-173-707-97
36 120 6.1 1894 12 US-10-174-569-97
37 120 6.1 1894 12 US-10-174-583-97
38 120 6.1 1894 12 US-10-174-587-97
39 120 6.1 1894 12 US-10-174-589-97
40 120 6.1 1894 12 US-10-174-591-97
41 120 6.1 1894 12 US-10-175-736-97
42 120 6.1 1894 12 US-10-175-742-97
43 120 6.1 1894 12 US-10-175-744-97
44 120 6.1 1894 12 US-10-175-745-97
45 120 6.1 1894 12 US-10-175-748-97

ALIGNMENTS

RESULT 1

US-09-073-009-138
Sequence 138, Application US/09073009
Patent No. US20010012888A1
GENERAL INFORMATION:
APPLICANT: Alderson, Mark
APPLICANT: Dillon, David C.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Campos-Neto, Antonio
TITLE OF INVENTION: COMPOUNDS AND DIAGNOSIS OF
NUMBER OF SEQUENCES: 144
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY
STREET: 6300 Columbia Center, 701 Fifth Ave.
CITY: Seattle
STATE: Washington
COUNTRY: US
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/073.009
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Mark, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.441C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-622-4900
TELEFAX: 206-682-6031
INFORMATION FOR SEQ ID NO: 138:
SEQUENCE CHARACTERISTICS:
LENGTH: 392 amino acids
TYPE: amino acid

TUBERCULOSIS AND ME

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-073-009-138

Query Match      100.0%; Score 1969; DB 9; Length 392;
Best Local Similarity 100.0%; Pred. No. 9.4e-166;
Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSRAFIIDPTISAIDGLYDILGIPNQGGILYSSLEYFEKALEELAAAFPGDGLGSA 60
Db 1 MSRAFIIDPTISAIDGLYDILGIPNQGGILYSSLEYFEKALEELAAAFPGDGLGSA 60

QY 61 DYAGKRNHNHVPFQELADLDROLISLIHQANAVQTTTRDILEGAKKGLFVRPVAVDLT 120
Db 61 DYAGKRNHNHVPFQELADLDROLISLIHQANAVQTTTRDILEGAKKGLFVRPVAVDLT 120

QY 121 YIPVVGHALSAAFPQAPFCAGAMAVVGGALAYLVVKTLLINATQLKLLAKLAELVAAAIAD 180
Db 121 YIPVVGHALSAAFPQAPFCAGAMAVVGGALAYLVVKTLLINATQLKLLAKLAELVAAAIAD 180

QY 181 IISDVADIIGKTLGEVWEFTTNALNGKELWDKLTGWVTGLFSRGSNLSLESPFAGVPGLT 240
Db 181 IISDVADIIGKTLGEVWEFTTNALNGKELWDKLTGWVTGLFSRGSNLSLESPFAGVPGLT 240

QY 241 GATSGLSQVTLGFGAAGLSASSGLAHADSLASSASLPALAGTGGSGFGGLPSLAQVHAA 300
Db 241 GATSGLSQVTLGFGAAGLSASSGLAHADSLASSASLPALAGTGGSGFGGLPSLAQVHAA 300

QY 301 STRQALRPRADGPGVAAAEEQVGSQSLVSAQSQGMPGVGMGHPSSGASKGTTTKKY 360
Db 301 STRQALRPRADGPGVAAAEEQVGSQSLVSAQSQGMPGVGMGHPSSGASKGTTTKKY 360

QY 361 SEGAAAGTDAERAPVEADAGGQKVLVRNV 392
Db 361 SEGAAAGTDAERAPVEADAGGQKVLVRNV 392

RESULT 2
US-09-793-306-138
; Sequence 138, Application US/09793306
; Patent No. US20020098200A1
; GENERAL INFORMATION:
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Skeiky, Yasir
; APPLICANT: Ovendale, Pamela
; APPLICANT: Jen, Shvian
; APPLICANT: Lodes, Michael
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compounds and Methods for Diagnosis and Immunotherapy
; FILE REFERENCE: 014058-008740US
; CURRENT APPLICATION NUMBER: US/09/793,306
; CURRENT FILING DATE: 2001-02-26
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: US 60/223,828
; NUMBER OF SEQ ID NOS: 164
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 138
; LENGTH: 392
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: hTcc#1
US-09-793-306-138

Query Match      100.0%; Score 1969; DB 9; Length 392;
Best Local Similarity 100.0%; Pred. No. 9.4e-166;
Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSRAFIIDPTISAIDGLYDILGIPNQGGILYSSLEYFEKALEELAAAFPGDGLGSA 60
Db 1 MSRAFIIDPTISAIDGLYDILGIPNQGGILYSSLEYFEKALEELAAAFPGDGLGSA 60

QY 61 DYAGKRNHNHVPFQELADLDROLISLIHQANAVQTTTRDILEGAKKGLFVRPVAVDLT 120
Db 61 DYAGKRNHNHVPFQELADLDROLISLIHQANAVQTTTRDILEGAKKGLFVRPVAVDLT 120

QY 121 YIPVVGHALSAAFPQAPFCAGAMAVVGGALAYLVVKTLLINATQLKLLAKLAELVAAAIAD 180
Db 121 YIPVVGHALSAAFPQAPFCAGAMAVVGGALAYLVVKTLLINATQLKLLAKLAELVAAAIAD 180

QY 181 IISDVADIIGKTLGEVWEFTTNALNGKELWDKLTGWVTGLFSRGSNLSLESPFAGVPGLT 240
Db 181 IISDVADIIGKTLGEVWEFTTNALNGKELWDKLTGWVTGLFSRGSNLSLESPFAGVPGLT 240

QY 241 GATSGLSQVTLGFGAAGLSASSGLAHADSLASSASLPALAGTGGSGFGGLPSLAQVHAA 300
Db 241 GATSGLSQVTLGFGAAGLSASSGLAHADSLASSASLPALAGTGGSGFGGLPSLAQVHAA 300

QY 301 STRQALRPRADGPGVAAAEEQVGSQSLVSAQSQGMPGVGMGHPSSGASKGTTTKKY 360
Db 301 STRQALRPRADGPGVAAAEEQVGSQSLVSAQSQGMPGVGMGHPSSGASKGTTTKKY 360

QY 361 SEGAAAGTDAERAPVEADAGGQKVLVRNV 392
Db 361 SEGAAAGTDAERAPVEADAGGQKVLVRNV 392

RESULT 3
US-10-098-732A-29
; Sequence 29, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Gudarian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; CURRENT FILING DATE: 2003-04-29
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 29
; LENGTH: 392
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: MTB40 (HTCC#1)
US-10-098-732A-29

Query Match      100.0%; Score 1969; DB 12; Length 392;
Best Local Similarity 100.0%; Pred. No. 9.4e-166;
Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSRAFIIDPTISAIDGLYDILGIPNQGGILYSSLEYFEKALEELAAAFPGDGLGSA 60
Db 1 MSRAFIIDPTISAIDGLYDILGIPNQGGILYSSLEYFEKALEELAAAFPGDGLGSA 60

QY 61 DYAGKRNHNHVPFQELADLDROLISLIHQANAVQTTTRDILEGAKKGLFVRPVAVDLT 120
Db 61 DYAGKRNHNHVPFQELADLDROLISLIHQANAVQTTTRDILEGAKKGLFVRPVAVDLT 120

QY 121 YIPVVGHALSAAFPQAPFCAGAMAVVGGALAYLVVKTLLINATQLKLLAKLAELVAAAIAD 180
Db 121 YIPVVGHALSAAFPQAPFCAGAMAVVGGALAYLVVKTLLINATQLKLLAKLAELVAAAIAD 180

QY 181 IISDVADIIGKTLGEVWEFTTNALNGKELWDKLTGWVTGLFSRGSNLSLESPFAGVPGLT 240
Db 181 IISDVADIIGKTLGEVWEFTTNALNGKELWDKLTGWVTGLFSRGSNLSLESPFAGVPGLT 240

```


QY 241 GATSGLSQVTLFGAAGLSASSGLAHADSLASSASLPALAGIGGSGFGGLPSLAOVHAA 300
Db 241 GATSGLSQVTLFGAAGLSASSGLAHADSLASSASLPALAGIGGSGFGGLPSLAOVHAA 300
QY 301 STRQALRPADGPVGAABEQVGSQSLVSAQSGQMGPGVGMGMHPSSGASKGTTTKY 360
Db 301 STRQALRPADGPVGAABEQVGSQSLVSAQSGQMGPGVGMGMHPSSGASKGTTTKY 360
QY 361 SEGAAAGTDAERAPVEADAGGQKVLVRNV 392
Db 361 SEGAAAGTDAERAPVEADAGGQKVLVRNV 392

RESULT 4

US-10-080-170-610
; Sequence 610, Application US/10080170
; Publication No. US20030129601A1
; GENERAL INFORMATION:
; APPLICANT: COLE, S.T.
; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR
; TITLE OF INVENTION: IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR
; FILE REFERENCE: 03495.0218
; CURRENT APPLICATION NUMBER: US/10/080,170
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: 60/270,123
; PRIOR FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 652
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 610
; LENGTH: 392
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-080-170-610

Query Match 100.0%; Score 1969; DB 16; Length 392;
Best Local Similarity 100.0%; Pred. No. 9.4e-166;
Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSRAFIIDPTISAIDGLYDLGIGIPNQGILYSSLEYFEKALELAAFPDGDGLGSA 60
Db 1 MSRAFIIDPTISAIDGLYDLGIGIPNQGILYSSLEYFEKALELAAFPDGDGLGSA 60
QY 61 DKYAGKRNHVNFFQELADLDQLISLIHQANAVOTTRDILEGAKKGLFVRPVAVDLT 120
Db 61 DKYAGKRNHVNFFQELADLDQLISLIHQANAVOTTRDILEGAKKGLFVRPVAVDLT 120
QY 121 YIPVVGHALSAAPFCAGAMAVVGALAYLVVKTINATQLKLLAKLAELVAAIAD 180
Db 121 YIPVVGHALSAAPFCAGAMAVVGALAYLVVKTINATQLKLLAKLAELVAAIAD 180
QY 181 IISDVADIKTGLGEVWEFITNALNGLKELWDLKLTGWVTGLFSRGWSNLESFFAGVPG 240
Db 181 IISDVADIKTGLGEVWEFITNALNGLKELWDLKLTGWVTGLFSRGWSNLESFFAGVPG 240
QY 241 GATSGLSQVTLFGAAGLSASSGLAHADSLASSASLPALAGIGGSGFGGLPSLAOVHAA 300
Db 241 GATSGLSQVTLFGAAGLSASSGLAHADSLASSASLPALAGIGGSGFGGLPSLAOVHAA 300
QY 301 STRQALRPADGPVGAABEQVGSQSLVSAQSGQMGPGVGMGMHPSSGASKGTTTKY 360
Db 301 STRQALRPADGPVGAABEQVGSQSLVSAQSGQMGPGVGMGMHPSSGASKGTTTKY 360
QY 361 SEGAAAGTDAERAPVEADAGGQKVLVRNV 392
Db 361 SEGAAAGTDAERAPVEADAGGQKVLVRNV 392

RESULT 5

US-10-080-170-53
; Sequence 53, Application US/10080170
; Publication No. US20030129601A1

; GENERAL INFORMATION:
; APPLICANT: COLE, S.T.
; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR
; TITLE OF INVENTION: IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR
; FILE REFERENCE: 03495.0218
; CURRENT APPLICATION NUMBER: US/10/080,170
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: 60/270,123
; PRIOR FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 652
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 53
; LENGTH: 394
; TYPE: PRT
; ORGANISM: Mycobacterium leprae
US-10-080-170-53

Query Match 63.9%; Score 1258; DB 16; Length 394;
Best Local Similarity 62.7%; Pred. No. 7e-103;
Matches 247; Conservative 55; Mismatches 90; Indels 2; Gaps 1;
QY 1 MSRAFIIDPTISAIDGLYDLGIGIPNQGILYSSLEYFEKALELAAFPDGDGLGSA 60
Db 1 MSRAFIIDPTISAIDGLYDLGIGIPNQGILYSSLEYFEKALELAAFPDGDGLGSA 60
QY 61 DKYAGKRNHVNFFQELADLDQLISLIHQANAVOTTRDILEGAKKGLFVRPVAVDLT 120
Db 61 DKYAGKRNHVNFFQELADLDQLISLIHQANAVOTTRDILEGAKKGLFVRPVAVDLT 120
QY 121 YIPVVGHALSAAPFCAGAMAVVGALAYLVVKTINATQLKLLAKLAELVAAIAD 180
Db 121 YIPVVGHALSAAPFCAGAMAVVGALAYLVVKTINATQLKLLAKLAELVAAIAD 180
QY 181 IISDVADIKTGLGEVWEFITNALNGLKELWDLKLTGWVTGLFSRGWSNLESFFAGVPG 240
Db 181 IISDVADIKTGLGEVWEFITNALNGLKELWDLKLTGWVTGLFSRGWSNLESFFAGVPG 240
QY 241 GATSGLSQVTLFGAAGLSASSGLAHADSLASSASLPALAGIGGSGFGGLPSLAOVHAA 300
Db 241 GATSGLSQVTLFGAAGLSASSGLAHADSLASSASLPALAGIGGSGFGGLPSLAOVHAA 300
QY 301 STRQALRPADGPVGAABEQVGSQSLVSAQSGQMGPGVGMGMHPSSGASKGTTTK 358
Db 301 STRQALRPADGPVGAABEQVGSQSLVSAQSGQMGPGVGMGMHPSSGASKGTTTK 358
QY 359 KYSEGAAGTDAERAPVEADAGGQKVLVRNV 392
Db 359 KYSEGAAGTDAERAPVEADAGGQKVLVRNV 392
QY 361 KYSEGAAGTDAERAPVEADAGGQKVLVRNV 394
Db 361 KYSEGAAGTDAERAPVEADAGGQKVLVRNV 394

RESULT 6

US-09-820-843A-18
; Sequence 18, Application US/09820843A
; Publication No. US20030039963A1
; GENERAL INFORMATION:
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEI
; FILE REFERENCE: Q63915
; CURRENT APPLICATION NUMBER: US/09/820,843A
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 18
; LENGTH: 256
; TYPE: PRT
; ORGANISM: M. tuberculosis
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: FE_PERS
; NAME/KEY: misc feature
; OTHER INFORMATION: gi|3261822

```
US-09-820-843A-18
Query Match      7.1%; Score 140.5; DB 11; Length 256;
Best Local Similarity 30.6%; Pred. No. 0.0025;
Matches 57; Conservative 11; Mismatches 77; Indels 41; Gaps 7;

QY 219 TGLFRGWSNLSFPAGVPLTGTATSGLSQVTLFGAAGLASSGLAHADSLASSA-SLP.277
Db 82 TCGFAQGG---RGGFGG-NGNTGASGGMG---GAGGAGGAGGAGLLIGDGGAGGAGG 134
QY 278 ALAGIGGGSGFGLPSLAQVHA-----ASTQALRPRADGPVGA 318
Db 135 GAGVGGGGAGGTGGGGVASFPGGNAGFGRRGGDGGDGGTGGAGGAGGAGGAG 194
QY 319 EQVGQSQVLSAQSGQGMGPVGMGMHPSGASGKTGTTTKYSEGAAGTDAERAPVEA 378
Db 195 GWLSGSHGAHGAAGSGGGGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 240
QY 379 DAGGGQ 384
Db 241 GAPGTQ 246

RESULT 7
US-10-259-678-727
; Sequence 727, Application US/10259678
; Publication No. US20030198974A1
; GENERAL INFORMATION:
; APPLICANT: Cole, Stewart
; APPLICANT: Buchrieser-Brosch, Roland
; APPLICANT: Gordon, Stephen
; APPLICANT: Billaud, Alain
; TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM
; TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
; FILE REFERENCE: 3495-0169
; CURRENT APPLICATION NUMBER: US/10/259,678
; PRIOR FILING DATE: 2002-09-30
; PRIOR FILING DATE: 1998-04-16
; NUMBER OF SEQ ID NOS: 743
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 727
; LENGTH: 318
; TYPE: PRT
; ORGANISM: Mycobacterium sp.
US-10-259-678-727

Query Match      6.9%; Score 135; DB 12; Length 318;
Best Local Similarity 27.5%; Pred. No. 0.001;
Matches 58; Conservative 11; Mismatches 104; Indels 38; Gaps 7;

QY 191 GTLGEVWEFTNALNGLKELMDKLTGWTGLFSRGWSNLSFPAGVPLTGTATSGLSQVT 250
Db 14 GTPCAGVSGATGAPG-----GWLIDGGAGGSGAG--SGAPGAGGAGAGLWGTG 61
QY 251 GLFCAAGLASSG-----LAHADSLASSASLPALAGIGGGSGFGGLPSLAQVH 298
Db 62 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 121
QY 299 AASTROALRPRADGPVGAARQVGGQSLVSAGGSG-----MGSPVGMGMHPSGAS 352
Db 122 GAGGTGLV--GGDGGAGGAGGTGGLLAGLIGAGGGGTGGGLSTNGDGGVGG-----AGN 175
QY 353 KGTITTKYSEGAAGTDAERAPVEADAGG 383
Db 176 AGMLAGPGGAGGAGG--DGNLDTGGDGGAG 204

RESULT 8
US-09-964-662-1
; Sequence 1, Application US/09964662
; Publication No. US20030166846A1
; GENERAL INFORMATION:
; APPLICANT: PROTEIN SPECIALTIES LTD.
; APPLICANT: HSC RESEARCH AND DEVELOPMENT LIMITED PARTNERSHIP
; TITLE OF INVENTION: SELF-ALIGNING PEPTIDES MODELED ON HUMAN ELASTIN AND
; TITLE OF INVENTION: OTHER FIBROUS PROTEINS
; FILE REFERENCE: 041082/0112
; CURRENT APPLICATION NUMBER: US/09/964,662
; CURRENT FILING DATE: 2003-05-08
; PRIOR APPLICATION NUMBER: 09/340,736
; PRIOR FILING DATE: 1999-06-29
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 1
; LENGTH: 731
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-964-662-1

Query Match      6.8%; Score 133.5; DB 12; Length 731;
Best Local Similarity 20.4%; Pred. No. 0.0046;
Matches 86; Conservative 35; Mismatches 136; Indels 165; Gaps 15;

QY 22 GIGIPN--QGILYSSLEYFEKALELAAAFDGDGHLGSAADKYACKNHNHVFQELAD 79
Db 117 GVGLPGVYPGVVL-----PGARFPGVGLPGV----- 143
QY 80 LDRQLSLIHDOANAVOTTRDILEGAKKGLFVRPV-----AVDLTYIPVVGHALSAAPQ 134
Db 144 -----PTGAGVKPKAPGVGGAGFAGIPGVPGPGPQGVPLGY-PIKAPKLPGGYG 192
QY 135 APFCAGAM-----AVWGA-LAYLVVKTINATOLLKLLAKLAEVAAAIADISDA 186
Db 193 LPYTTGKLPGYGGPVGAGAGKAGYPTGTGVPQAAAAAATAAKAFGAGAGVLPVGV 252
QY 187 DIIKGLTGEWEFTNALNGLKELMDKLTGWTGLFSRGWSNLSFPAGVPLTGTATSGL 246
Db 253 G-----AQVHAASTROALRPRADGPVGAARQVGG-----AGVPGVGA----- 262
QY 247 SQVTGLFGAAGLASSGLAHADSLASSASLPALAG-IGGSGSGF-----GLPSL--- 294
Db 263 --IPGIGGIAGVGTAAAAAATAAKAAGAGAGVPGVPGVPGVPGVPGVPGVPGV 320
QY 295 -----AQVHAASTROALRPRADGPVGAARQVGG-----AGVPGVGA----- 322
Db 321 PGAGIPVPGAGIPGAAVPGVSPAAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAA 380
QY 323 ---GQSOLVSAQSGQGMGFGVGMG-----GMHPSGASGKTGTTTKYSEG---AAAGTDAE 372
Db 381 FGVGVGIPGVAGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGV 440
QY 373 RA 374
Db 441 KA 442

RESULT 9
US-10-156-761-13993
; Sequence 13993, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
```

PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 13993
LENGTH: 446
TYPE: PRT
ORGANISM: Streptomyces avermitilis
10-156-761-13993

```

Query Match      6.7%; Score 131.5; DB 15; Length 446;
Best Local Similarity 25.7%; Pred. No. 0.0034;
Matches 74; Conservative 113; Mismatches 112; Indels 69; Gaps 11;

94  AVQTTDILEGAKGLEFVRVAVDLYTIPVVGHALSAAPQAPFCA---GMAVVGGLAL- 149
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
51  AITSGLALLEVRDGLGSGSVAGLTSVPEPLCFVFGV-MAPRLARRRGAAVVGAMV 109
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
150 ---AYLVKXTLNATQLKLAKLAELVAAADIISDVADIKGTGLVEVFEITNALG 206
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
110 AITAGLVIRPYAGGTAGF-LAASALAMGVAVSNLMPVI----- 148
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
207 LKELMDKLTGWYTGFLSRGWNLESFFAGVPGITGATSGLSQVT-GLFGAAGLSASSGLA 265
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
149 VKRWFPDRVGSMTGLYSM-----ALALGTSAAAAVTPLTDALGSGWQSGLA 195
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
266 HADSLASSASLP-----ALAGIGGSGFGGLSLACQVHAASTQALRPADCPVGAAAE 319
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
196 VWAGLAAAVLFWPIELVIRAGGAGSGHFGSSSAGSGSHLGSSEA-----AAAAAP 247
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
320 QVWGQSGLVSAQ-----GSGCMGPGVVMGMG---HPSGAGKGT 355
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
248 QEADGLIRITRSTAWALAVFFGLOATAAYITGCMWMOIIPRDAGVAGGT 295
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 10
 US-09-820-843A-20
 ; Sequence 20, Application US/09820843A
 ; Publication No. US20030039963A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Council of Scientific and Industrial Research
 ; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEIN
 ; TITLE OF INVENTION: USEFUL AS ANTI-INFECTIVES
 ; FILE REFERENCES: Q63915
 ; CURRENT APPLICATION NUMBER: US/09/820,843A
 ; CURRENT FILING DATE: 2001-03-30
 ; NUMBER OF SEQ ID NOS: 118
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 20
 ; LENGTH: 1079
 ; TYPE: PRT
 ; ORGANISM: M. tuberculosis
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; OTHER INFORMATION: PE_FGRS
 ; NAME/KEY: misc feature
 ; OTHER INFORMATION: gi|2924449
 US-09-820-843A-20

```

Query Match.          6.6%; Score 130; DB 11; Length 1079;
Best Local Similarity 26.4%; Pred. No. 0.016;
Matches 46; Conservative 12; Mismatches 72; Indels 44; Gaps 4;

QY 234 AGVPLGTGATSGLSQVVTGLFCAAGLSASSGLAHADSLASSASLPA----- 278
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 313 AGAGGAAGSNGGTVGANGTGDDGNGGGAAGAATAGSNGGAGTGAGGNGGTGGRGSGGA 372
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 279 ----LAGIGGSGFGGLPSLAQVHAASTQALRPADGPGVCAAAEQVGQSQLVASGQS 334
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 373 GGDGIGVGVGKKGNG-----ADGEVGGAG-GAGGSGPNTSPCGNG 412
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 335 GMGGPVMGMNHPSGSA-----SKGTTTKYSEGAAGTDAERAPVEADAGGG 383
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 413 GGGGQGGSGAGGAAGAGGAGGANGTAGNGCGGAGGTGCGAGAASATNGGSG 466
      : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 11
US-09-961-403-8
; Sequence 8, Application US/09961403
; Publication No. US20030077589A1
; GENERAL INFORMATION:
; APPLICANT: HE-STUMPP, HOLGER
; APPLICANT: HAENDLER, BERNARD
; APPLICANT: KRAETZSCHMAR, JOERN
; APPLICANT: KREFT, BERTHOLT
; APPLICANT: WINTERHAGER, ELKE
; APPLICANT: REGIDOR, PEDRO
; APPLICANT: SCOTTI, SIMONE
; TITLE OF INVENTION: METHOD FOR IN VITRO DIAGNOSIS OF ENDOMETRIOSIS
; FILE REFERENCE: SCH-1789
; CURRENT APPLICATION NUMBER: US/09/961.403
; CURRENT FILING DATE: 2001-09-25
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 730
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-961-403-8

Query Match      6.6%; Score 129.5; DB 11; Length 730;
Best Local Similarity 20.3%; Pred. No. 0.01;
Matches      86; Conservative      36; Mismatches      134; Indels      167; Gaps      15

Qy      22  GTGIPN--GGGILYSSLEYFEKALEELAAAPPDGMGLGSAAADKYAGKNRNHNVNFFOELAD 79
Db      143  GVGLPGVPGGVL-----PCARPPGVGVLPGV-----169

Qy      80  LDRQLISLIHQANAVQTRDILBGAKKGLFEVPPV-----AYDLTVIPVGHALSAATQ 134
Db      170  -----PTGACVKPKACGVGAGTAGIPGVGPGFGPGQPPVPLGY-PIKAPKLPGGYG 218

Qy      135  APFCAGAM-----AVVGGG-LAYLVVYKTLINATQLKLKLAELIVAAAIADIISDVA 186
Db      219  LPVTTGKLPYGVGPGVAGAGKAGYPTGTGVPQAAAAAAXAKAAGAGVLPGVG 278

Qy      187  DIIKGTGLGVWEFIIINALNGKLKELWDKLITGVNVTGLFSRGWNSLESFFAGVGLTGTATSL 246
Db      279  G-----AGVPGVPGA-----288

Qy      247  SSVTLGLFAAGLSASSGLAHADSLASSASLPALAG-IGGGSGFG-----GLPSL-- 294
Db      289  --IPGIGTAGVGTAAAAAAXAKAAGAGLVPGFGFGPGVGVGVPNGVPGVGVG 346

Qy      295  -----AQVHAASTQALRPACGPGVCAAAEQVG-----322
Db      347  PGAGIPVVPVPGAGIFGAAVPGVVSPEAAAKAAXAKAAGYARPPGVGVGGIPTTYGVGAGGFGP 406

Qy      323  -----GOSQLVSACSGMGCGPYGVGMHPSSGASKGTTTKYSGS--AAAGTEDA 371
Db      407  FGVGVGGIPGVAGVFSGVGVGVGVGV-GISFEAQAAAAAAXAKAAGYIGVTTPAAAAAKAA 465

Qy      372  ERA 374
Db      466  AKA 468

```

```

RESULT 12
US-10-097-111-283
; Sequence 283, Application US/10097111
; Publication No. US20030138771A1
; GENERAL INFORMATION:
; APPLICANT: PELLETIER, JERRY
; APPLICANT: GROS, PHILIPPE
; APPLICANT: DUBOW, MICHAEL
; TITLE OF INVENTION: DNA SEQUENCES FROM S. PNEUMONIAE BACTERIOPHAGE EPI THAT
; TITLE OF INVENTION: ENCODE ANTI-MICROBIAL POLYPEPTIDES

```

```

; FILE REFERENCE: 073406-0603
; CURRENT APPLICATION NUMBER: US/10/097,111
; PRIOR FILING DATE: 2002-07-24
; PRIOR APPLICATION NUMBER: 09/676,412
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/157,218
; PRIOR FILING DATE: 1998-09-30
; NUMBER OF SEQ ID NOS: 552
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 283
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-097-111-283

Query Match      6.6%; Score 129; DB 12; Length 1149;
Best Local Similarity 19.8%; Pred. No. 0.022;
Matches 87; Conservative 64; Mismatches 150; Indels 138; Gaps 16;

QY      6 IIDPTISAIDGLYDLIGIPNQGGILYSSLEYFEKALEELAAAPPDGMGLGSAADKYAG 65
Db      686 IIAAVQIITALENGVLQALPT---LIQAGLILSALINGLVQALPA-----729

QY      66 KNRNVNFFQELADLRQII-SLIHQANAVOTTRDILLEGAKKGLFVRPVAVDLTYPV 124
Db      730 -----IIQAAVQIIMSIVQALIEIENLPMIIEAAMQIIMGLVNAL-----IEN 770

QY      125 VGHLSAAQAPFCAGAMAVGALAYL-----VVKTLI-----NATOLLK-----165
Db      771 IGPILQAGI-----LMALIEGLIQVLPBELTAIQTISLLEAILSNLPOLLEAGVKL 825

QY      166 -----LLAKLAELVAAAI-----ADIISDVAD-----IIGTLEGEVWEFI 200
Db      826 LLSLLOGLNMLPOLITAGALQIMMALLKAVIDFVPKLLQAGVOLLKALIQIASIILGSL 885

QY      201 TNALNGLKELDKLTGWTLGFSRGWSNLESFFAGVPGLTGA-----TSGLSOVTG 251
Db      886 STAGNMLSLVSKIASFVGQMVSGGANLRNFISGIGMIGSAVSKISMGTSIVSKVIG 945

QY      252 LFG---AAGLSASSGLAHADSLASSASLPALAGIGGGS-----GFGGLPSLAQVHAAS 303
Db      946 FAGQMVYAGVNLVRGFIINGISSVSSAVSAANASSALNAVKGFLGIHSPRV-----999

QY      304 QALRPADGPGVAAAPQVGGQQLVSAQSGQMGPGVGMGMHPSGSKAGTITTKYSEG 363
Db      1000 -----MEQMGIIY-----GQGFVNGIGNM-----ITTRDKAKEM 1029

QY      364 AAAGTDAERAPVEADAGG*382
Db      1030 AETVTEALSDVRNDIQENG 1048

RESULT 13
US-09-738-626-6032
; Sequence 6032, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR FILING DATE: 2000-12-18
; PRIOR FILING DATE: 1999-12-16

```

```

; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: Patentin ver. 3.0
; SEQ ID NO 6032
; LENGTH: 585
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6032

Query Match      6.4%; Score 127; DB 10; Length 585;
Best Local Similarity 21.0%; Pred. No. 0.013;
Matches 115; Conservative 54; Mismatches 186; Indels 192; Gaps 24;

QY      12 SAIDGLYDLIGIG-----IPNOGILYSSLEYFEKALEELAAAPPDGMGLGSAADKYAGKN 67
Db      41 SPVSGGL-DQLSGHGNVINGGAGANVLSYAEQIEWLSAALQASGAALTQDDELFAFG 99

QY      68 RHNWNPFOELADLRQLIS-----LIHQANAVOTTRDIL-----EGAKK 107
Db      100 -----MDVADTGRVVEESVMFPARPAPRPFESFVFNPPAVSPSLDALCSQSGTNS 152

QY      108 G-----LEFVRPVAVDLTYPVVGHALS-----AAFOAPE 138
Db      153 GAVLEAQSGWSMASAISNVASLS--SIAGEILAEISGETFEQAAARINEVAAGATPA 210

QY      139 AGAM---AVVGGG-----LAYLVVKTLLINATQLKL-LAKLAELVAAAIADI 182
Db      211 ANAKMGCASVGTLRNRYMGHRQVFWAATSIRAILDPVQLAAERAFASFOATFQADVL 270

QY      183 SDVADI-----IKGTLEGEVWEFI--TNALNGLKELWDLTGWTLGFSRGWSNLESFF--233
Db      271 TGMPPVSNLMQMGANGSAGEIALGMDIAGSQAWSAAGLTFSGAAQGGVANAGSIAPD 330

QY      234 AGVPGLTGATSGLSQVGTGLFG-----AAGLSASSGLAHADSLASSA 274
Db      331 AAVQGAAGQSG-----VGSFETVTDQLDGINIGDMLTSAASAGOSLANGLAMPTSPNSA 385

QY      275 S--LPA-----LAGIGGSGFG-----GLP 292
Db      386 SGAIPSSMSAASPLGAFGSGAGLGAQGGISGAPGAISRAAGSAGSGVPGMTGGPGAP 445

QY      293 SL-----AQVHAASRQALRP-----RADGPVGA-----AAEQVGQSQ 326
Db      446 GITSDSLNGARTHGASSAGAVAPMGGAGMSGVVVGAGTGSQSKYARQTGSSVSSSQ 505

QY      327 LVSAQGSQGMG-----GPFVGMGMHPSGSKG-TTTTKYSEGAAGATEDAERA 374
Db      506 SGSGLGMVSGSGKPSISNFCRGMMPMPMPMGAGGAGGQKNTGKVTVTSAVEDRNLA 565

QY      375 PVEADAG 381
Db      566 ALLGDRG 572

RESULT 14
US-10-259-678-728
; Sequence 728, Application US/10259678
; Publication No. US20030198974A1
; GENERAL INFORMATION:
; APPLICANT: Cole, Stewart
; APPLICANT: Buchrieser-Brosch, Roland
; APPLICANT: Gordon, Stephen
; APPLICANT: Billault, Alain
; TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM
; TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
; TITLE OF INVENTION: LIBRARY APPLICATION TO THE DETECTION OF MYCOBACTERIA
; FILE REFERENCE: 3495-0169
; CURRENT APPLICATION NUMBER: US/10/259,678
; CURRENT FILING DATE: 2002-09-30
; PRIOR APPLICATION NUMBER: US/09/060,756

```

```

; PRIOR FILING DATE: 1998-04-16
; NUMBER OF SEQ ID NOS: 743
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 728
; LENGTH: 334
; TYPE: PR1
; ORGANISM: Mycobacterium sp.
US-10-259-678-728

```

Query Match	6.4%	Score 126.5	DB 12	Length 334
Best Local Similarity	27.6%	Pred. No. 0.0063		
Matches	55	Conservative 11	Mismatches 90	Indels 43
				Gaps 8
QY	191	GTLGWVWFETNALNGLKELWDLTGMVTGLFSGRGSNLSFAGVPGILGTGATSGLSQVT	250	
Db	14	GTFGAVGSGATCAPG-----GWLDDGGAGGSGAAG--SGAPGGAGGAAGL---W	58	
QY	251	GLFGAAGTJUSGSLAHADSLASSALPALAGICGGSGFGGLPSLAQWHAASTQALRPRA	310	
Db	59	GTGGAGGIGGAS-----TVLGTTGGGGVGGGLMGAGGAGGAGGTGLV--GG	102	
QY	311	DGPVCAAAEYQGGOSQLVSAGGSG-----MGPVQMGGMHPSSCASGKTTTKYISGGA	364	
Db	103	DGAGGAGGTGGLIAGLTIGAGGGHGTGTGGLSTNGDGGVGG---AGGNAGMLAGPCGAGG	158	
QY	365	AAGTEDAERAPVEADAGG	383	
Db	159	AGG--DGENLDTGGGGAG	175	

```

RESULT 15
US-10-156-761-11729
/ Sequence 11729, Application US/10156761
/ Publication No. US20030119018A1
/ GENERAL INFORMATION:
/ APPLICANT: OMURA, SATOSHI
/ APPLICANT: IKEDA, HARUO
/ APPLICANT: ISHIKAWA, JUN
/ APPLICANT: HORIKAWA, HIROSHI
/ APPLICANT: SHIBA, TADAYOSHI
/ APPLICANT: SAKAKI, YOSHIYUKI
/ APPLICANT: HATTORI, MASAHIRA
/ TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
/ FILE REFERENCE: 249-262
/ CURRENT APPLICATION NUMBER: US/10/156,761
/ CURRENT FILING DATE: 2002-05-29
/ PRIOR APPLICATION NUMBER: JP 2001-204089
/ PRIOR FILING DATE: 2001-05-30
/ PRIOR APPLICATION NUMBER: JP 2001-272697
/ PRIOR FILING DATE: 2001-08-02
/ NUMBER OF SEQ ID NOS: 15109
/ SEQ ID NO 11729
/ LENGTH: 749
/ TYPE: PRT
/ ORGANISM: Streptomyces avermitilis
US-10-156-761-11729

```

[illegible]

```

QY 156 TLINATQLLKLAELVAAAAIADIIISDVADIINKGTIGEVMFEFTINA----- 203
Db 398 FMAGIFHOYTTVA-----LAPYIAALIGMGTAVLWEERTKYWASLTLAGAVVASVMGVV 452
QY 204 -LNGLKEL-----WDKLTGWVTGLFSRGSWNSLESFFACVPG--LTGATSGLSQVTLGFGA 255
Db 453 LLNRTSDYLPWLKWFVLVGGHAGA-----GLVFAAKLGRRLAALAVGLSVFVASVADP 505
QY 256 AGLSASS-GLAHASDASSASLPALAGI-----GGSGSF-GGLPSLAQVH-----AAST 302
Db 506 TAYTLSTVNTGHTSGIVTAG--PAGAMGGRGGPGGGGMRGGFPAGQNQQGGTGN 563
QY 303 QOALRPADGPPVGAAGEQ-----VGGQSOLVSAQSQ-----GNGG 338
Db 564 QQGGPGGGQGGGGMGPPTGGFGGNQNGNGNAGQNGQNGMPPGGGTGERGGMGG 623
QY 339 PVNGMGHPPSGASKGTTTKYISE-----GAAAGTEDA 371
Db 624 --GNGGL--LISGASVSAAKKLBTEDADSYTWAAAAAICAQNA 661

```

Search completed: November 5, 2003, 17:07:22
Job time : 47 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 5, 2003, 16:59:57 ; Search time 29 Seconds
(without alignments)
1299.935 Million cell updates/sec

Title: US-09-688-672a-14
Perfect score: 1969
Sequence: 1 MSRAFIIDPTISAIDGLYDL.....RAPVEADAGGQKVLVRNVV 392

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 76:.*
1: PIR1:.*
2: PIR2:.*
3: PIR3:.*
4: PIR4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1959	100.0	392	A70957	hypothetical prote
2	1258	63.9	394	S72814	hypothetical prote
3	380.5	19.3	402	E70656	hypothetical prote
4	188	8.5	496	H70839	hypothetical glyci
5	155	7.9	1329	E70917	hypothetical glyci
6	153	7.8	837	S30971	gene 26 protein -
7	152	7.7	801	F70824	hypothetical glyci
8	151	7.7	741	G70917	hypothetical glyci
9	150	7.6	1306	A70934	hypothetical glyci
10	149.5	7.6	916	E70954	gene 116 protein -
11	148.5	7.5	515	T03323	hypothetical glyci
12	148.5	7.5	749	A70812	hypothetical glyci
13	148	7.5	409	A70647	probable PPE prote
14	148	7.5	588	F70971	hypothetical glyci
15	148	7.5	957	D70835	hypothetical glyci
16	148	7.5	1860	A70869	hypothetical glyci
17	147	7.5	439	D70954	hypothetical glyci
18	146.5	7.4	837	E70835	hypothetical glyci
19	145.5	7.4	1538	H70846	hypothetical glyci
20	145	7.4	914	F70987	hypothetical glyci
21	144.5	7.3	539	A70899	probable PPE prote
22	144	7.3	1381	E70806	hypothetical glyci
23	142.5	7.2	525	D70878	hypothetical glyci
24	142.5	7.2	667	A70893	hypothetical glyci
25	142.5	7.2	882	F70812	hypothetical glyci
26	142	7.2	461	F70571	hypothetical glyci
27	142	7.2	778	F70963	hypothetical glyci
28	141.5	7.2	1489	D70807	hypothetical glyci
29	141	7.2	584	G70804	hypothetical glyci

30	141	7.2	837	2	H72802	minor tail subunit
31	140.5	7.1	256	2	A70514	hypothetical glyci
32	140.5	7.1	853	2	A70896	hypothetical glyci
33	139.5	7.1	543	2	F70726	hypothetical glyci
34	137.5	7.0	783	2	E70824	hypothetical glyci
35	137.5	7.0	1901	2	F70806	hypothetical glyci
36	136	6.9	358	2	H70857	probable PPE prote
37	136	6.9	618	2	A70989	hypothetical glyci
38	136	6.9	731	2	C70974	hypothetical glyci
39	135.5	6.9	457	2	H70820	hypothetical glyci
40	135	6.9	923	2	E70820	hypothetical glyci
41	135	6.9	1517	2	T13329	hypothetical prote
42	134.5	6.8	361	2	G70682	hypothetical glyci
43	134	6.8	513	2	D70836	probable PPE prote
44	134	6.8	606	2	H70816	hypothetical glyci
45	133.5	6.8	940	2	F84089	phage-related prot

ALIGNMENTS

RESULT 1
A70957
hypothetical protein Rv3616c - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: A70957
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: A70957
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-392 <COL>
A:Cross-references: GB:295557; GB:AL123456; NID:G3242276; PIDN:CAB08950.1; PID:e316833;
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: Rv3616c

Query Match	100.0%;	Score	1959;	DB	2;	Length	392;
Best Local Similarity	100.0%;	Pred. No.	8.9e-119;	Mismatches	0;	Indels	0;
Matches	392;	Conservative	0;	Indels	0;	Gaps	0;
Qy	1	MSRAFIIDPTISAIDGLYDL	IGIPNQGGILYSSLEYFEKALEELAAAPPDGLGSA	60			
Db	1	MSRAFIIDPTISAIDGLYDL	IGIPNQGGILYSSLEYFEKALEELAAAPPDGLGSA	60			
Qy	61	DYAGKRNHVNFPQELADLRQLISLIHQANAVQTRDILEGAKGLBFRVAVDLT	120				
Db	61	DYAGKRNHVNFPQELADLRQLISLIHQANAVQTRDILEGAKGLBFRVAVDLT	120				
Qy	121	YIPVVGHALSAAFOAPFCAGAMVGGALAYLVVKTLLINATOLLKLAELVAAAIAD	180				
Db	121	YIPVVGHALSAAFOAPFCAGAMVGGALAYLVVKTLLINATOLLKLAELVAAAIAD	180				
Qy	181	IISVDADIIKGTGEVWEFTINALNGKLKELDKLTGVTGTFGRGWSNLSFFAGVPGLT	240				
Db	181	IISVDADIIKGTGEVWEFTINALNGKLKELDKLTGVTGTFGRGWSNLSFFAGVPGLT	240				
Qy	241	GATSGLSQVTLFGAAGLSASSGLAHADSLASSASIPALAGIGGSGFGGLPSLAQVHAA	300				
Db	241	GATSGLSQVTLFGAAGLSASSGLAHADSLASSASIPALAGIGGSGFGGLPSLAQVHAA	300				
Qy	301	STQALPRADPGPVGAAGAEQVGSQQLVSAQSGMGMPVGMGMHPSSGASKGTTTKY	360				
Db	301	STQALPRADPGPVGAAGAEQVGSQQLVSAQSGMGMPVGMGMHPSSGASKGTTTKY	360				
Qy	361	SEGAAGCTEADRAPVEADAGGQKVLVRNVV	392				

Db 361 SEGAAAGTDAERAPVEADAGGQKVLVRNV 392

RESULT 2

hypothetical protein B1620.C2.213 - Mycobacterium leprae
C:Species: Mycobacterium leprae
C:Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 23-Mar-2001
C:Accession: S72814
R:Smith, D.R.; Robinson, K.
A:Submitted to the EMBL Data Library, November 1993
A:Description: Mycobacterium leprae cosmid B1620.
A:Reference number: S72584
A:Accession: S72814
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-394 <SMI>
A:Cross-references: EMBL:U00015; NID:g466931; PIDN:AAC43223.1; PID:g466935

Query Match 63.9%; Score 1258; DB 2; Length 394;
Best Local Similarity 62.7%; Pred. No. 2.7e-73;
Matches 247; Conservative 55; Mismatches 90; Indels 2; Gaps 1;

QY 1 MSRAFIIDPTISAIDGLYDIIIGIPNOGGIYSSLEYFEKALEBIAAFAFGDGLWLSAA 60
Db 1 MSGAFIIDPTLKAIEAWHALLGIVPNDGGVLYSSLFPEKALEHIAAFAFGDGLWLSAA 60
QY 61 DKYAGNRHNVFFOELADRLQLISLHDQANAVQTRDILEGAKKGLFVRPVAVDLT 120
Db 61 DKYAGNRHNVFFOELADRLQLISLHDQANAVQTRDILEGAKKGLFVRPVAVDLN 120
QY 121 YIPVVGHALSAAPFCAGAMAVVGGALAYLVKTLINATQLKLLAKLAEVAAAIAD 180
Db 121 YIPVVGHALSAAPFCAGAMAVVGGALAYLVKTLINATQLKLLAKLAEVAAAIAD 180
QY 181 IISDVADIILKIGLGEVWEITNALNGKLKELWKLGTWGLFSGRWSNLESPFAGVPGUT 240
Db 181 VVSDGVAIIGKIGLGEVWEITNALNGKLKELWKLGTWGLFSGRWSNLESPFAGVPGUT 240
QY 241 GATSGLSQVTLFGAAGLSASSGLAHADSLASSASIPALAGIGGSGFGGLPGLSLAQVHAA 300
Db 241 GATSGLSQVTLFGVPGLSAGSGLLSGSLSTENLPSLAGVAGLGLSLPQLAQVHAA 300
QY 301 STRQALRPADGVPVGAABQVGSQOLVSAQSGQMGVPMGMPSSGASK--GTTTK 358
Db 301 STRQALRPADGVPVGAABQVGSQOLVSAQSGQMGVPMGMPSSGASK--GTTTK 358
QY 359 KYSEGAAGTDAERAPVEADAGGQKVLVRNV 392
Db 361 KYSEGAAGTDAERAPVEADAGGQKVLVRNV 394

RESULT 3

hypothetical protein RV3864 - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: E70656
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: E70656
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-402 <COL>
A:Cross-references: GB:Z83864; GB:AL123456; NID:g3261687; PIDN:CA806237.1; PID:e301440;
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: RV3864

Query Match 19.3%; Score 380.5; DB 2; Length 402;
Best Local Similarity 30.1%; Pred. No. 3.6e-17;
Matches 129; Conservative 47; Mismatches 157; Indels 95; Gaps 13;

QY 10 TISAIDGLYDIIIGIPNOGGIYSSLEYFEKALEBIAAFAFGDGLWLSAAADKYAGKNEN 69
Db 11 TSFIWQQLLLGEGIPDPGDIPTGSSILFKQISDKWGLAIPGTNWIGQAAEAYLNQIA 70
QY 70 HYNFFOELADRLQLISLHDQANAVQTRDILEGAKKGLFVRPVAVDLTIPVVGHAL 129
Db 71 QQLRAQVMGDLKLTGNMISNQAKYVSDTRDLVRAMKMDIGYKVKCKGLEKIPILGLHW 130
QY 130 SAAFOAPFCAGAMAVVGGALAYLVKTLINATQLKLLAKLAEVAAAIADIISDVADI 189
Db 131 SWELAIPTMSGIANAVVGGALLVITINTLNATNL----- 164
QY 190 KGTIGEVWEITN-----ALNGKELWKLGT-W-----VTGLFSGRW----- 226
Db 165 RGLGRLEMLTLPKFPGLPGLPSLPDIIDGLWPPKLPDIPGLPDPFDFKWPPT 224
QY 227 -----SNLESP-----FAGVPGITGATSGLSQVTLF-GAAGL----- 258
Db 225 PGSLPFDLPSPFGPFGFEPFAIPGP-ALPGLPSIPNLFPGLPGLDGLLFGVDLGLK 283
QY 259 -SASSGLAHADSLASSASIPALAGIGGSGFGGLPGLSLAQVHAASTRQALRPADGVPVGA 317
Db 284 PTWTELAALPDFLGGFAGLPSL-GFGNLLSFASLPTVGQVATMGQLQQLVAAGGSPSQL 342
QY 318 AEQVGGSQLVSAQSGQMGVPMGMPSSGASKGTTTKYSEGAAGTDAERAPVE 377
Db 343 ASMGSOQAQIISQAQOG-----GQQAATLVSDK-----KEDEEVA-----EAERAPID 387
QY 378 ADAGGGQK 385
Db 388 AGTAASQR 395

RESULT 4

hypothetical glycine-rich protein Rv0109 - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: H70839
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: H70839
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-496 <COL>
A:Cross-references: GB:AL021926; GB:AL123456; NID:g3261620; PIDN:CAA17303.1; PID:g2909574
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: Rv0109
C:Superfamily: Phaseolus glycine-rich cell wall protein 1.8

Query Match 8.5%; Score 168; DB 2; Length 496;
Best Local Similarity 25.5%; Pred. No. 0.0018;
Matches 82; Conservative 25; Mismatches 113; Indels 102; Gaps 13;

QY 115 VAVDLTIPVVGHALS---AAFOAPFCAGAMAVVG----- 146
Db 11 VAAATHLAGISALSTANAAAAAP--TTALSAGADEVSVLIAALFEAVAYQVQALSAQ 68
QY 147 -----GALAYLVKTLINATQLKLLAKLAEVAAAIADIISDVADI 192
Db 69 ALAFHDQFQVQALNNGVACYAAET-ANATPLQAL-----QTVQNVLTIVNAPTQALLGR 122

QY 193 LGEVWEFITHALNGKELWDLKLTGWVTG-----LFSRGWSNLESEFFAGVPGLTGATSGLSQ 248
Db 123 -----PFIGANGALPN-----TGQDGGPGGLLFGNGN-----GSGSGVDQ 159
QY 249 VTGLFGAAGL--SASSSLAHADSLASSASLPA-----LAGIGSGSGFGLPLSLAQVHAAS 301
Db 160 AGNGGAGLILNGSGSGVGPAGSAGAGAGAGLILFNGGPGAGGIGITGCD-----214
QY 302 TQALRPADPGVGAABAEQVGGSQLVSAAGSQMGPGVGMGHPSPSGASKGTTTKTKYS 361
Db 215 -----GGPGGAGGNAIG-----LFGSGGTGGMGGVGMGGVGNAGNGGCTAGLFG 261
QY 362 EGAAAGTDEAERAPVEADAGG 383
Db 262 HGGAGGAGGIGSADGLGGGGG 283
RESULT 5
E70917
hypothetical glycine-rich protein Rv1450c - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: E70917
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID: 98295987; PMID: 9634230
A: Accession: E70917
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-1329 <COL>
A: Cross-references: GB:295844; GB:AL123456; NID:93250713; PIDN: CAB09271.1; PID: g2131046
A: Experimental source: strain H37RV
C: Genetics:
A: Gene: Rv1450c
C: Superfamily: collagen alpha 2(I) chain; fibrillar collagen carboxyl-terminal homology
Query Match 7.9%; Score 155; DB 2; Length 1329;
Best Local Similarity 26.3%; Pred. No. 0.039;
Matches 79; Conservative 27; Mismatches 132; Indels 62; Gaps 13;
QY 110 EFVRPVAVDLTIYPVVGHALSAAFAFPFCAGAMAVV---GGALAYLVWKTLLN-----159
Db 9 ETVAALALDVARI---GSSIGRANAA--AAGSTTVLAAGADEVSAIATLFGSHAREYQ 63
QY 160 --ATQLLKLLAKLAEVAAIAADIISDVADIKGTLGEVWEFITHALN-----GLKEL 210
Db 64 AISTQVAAPHRDFAOTLSAAVGSVGAEEA--TNAAPLATLEHNVNLNALNAPTQALLGRPLI 122
QY 211 WDKL-----TGWVTGLFSRGWSNLESEFFAGVPELTGATSGLSQVTLFGAAGLSASSGLA 265
Db 123 GDGAAGAPGTGQAGGAGGILNGNGAGGAPQVGGAGG---AAGLFTGTGAGAGGAG 179
QY 266 HADSLASSASLPALAGIGGSGFGGLPLSLAQVHAASLTQALRPADPGVGAABAEQVGGOS 325
Db 180 AAGCAGGSGGW--LLNGGVGGAGG-----QSL-----LGGATGGAGGNA 217
QY 326 QLVSAGSQSGMGPGVGMGHPSPSGAS--KGITTKYSAGAAAGTDEAERAPVEADAGG 383
Db 218 GLFVGVTGGPGGPGGPGVGGTGGAGGLGGLTYAGGHHGAGGP-----GPIGVGGHG 272
RESULT 6
S30971
gene 26 protein - Mycobacterium phage L5
C: Species: Mycobacterium phage L5
C: Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 20-Jun-2000
C: Accession: S30971
R: Donnelly-Wu, M.K.; Jacobs Jr., W.R.; Hatfull, G.F.
Mol. Microbiol. 7, 407-417, 1993

A: Title: Superinfection immunity of mycobacteriophage L5: applications for genetic trans
A: Reference number: S30949; MUID: 93211283; PMID: 8459767
A: Accession: S30971
A: Status: nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-837 <DON>
A: Cross-references: EMBL: Z18946; NID: g15859; PIDN: CAA79402.1; PID: g579124
A: Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1992
C: Genetics:
A: Gene: 26
A: Start codon: GTG
Query Match 7.8%; Score 153; DB 2; Length 837;
Best Local Similarity 22.3%; Pred. No. 0.03;
Matches 101; Conservative 59; Mismatches 157; Indels 136; Gaps 18;
QY 13 AIDGLYDLIGIPNQGILYSSLEYPEKALEELAAAPFGDGLGSAADKVGAKRNHVN 72
Db 310 AMQGLSQTILG-----SVLNFNRLMESGLQAM---GQLG-----GFLSTFIN 348
QY 73 PFQE-----LADDRQLIISLHDOANAVQTRDILEGAKXG-LEFV 112
Db 349 GFGDLFVSLMPALTSVSLIGNVLGTQLAPIVTALTAPFQTLASTLGTMLTALQAL 408
QY 113 RPVAVDLTVI--PVVGHALSAAFAQ-----PFCAGAMAV 144
Db 409 GPI-----LIQVATLGTITLNTALQALQPMPLSMOSFQOISDLVLTSLAPHIPALATLQG 465
QY 145 VGGA---LAYLVKTLINATQLLKLAEL-----V 174
Db 466 VAGAVLQAPTIISTLTPA--FVLVPKVAELFTIYNLVQSPANLMPVLPPLAALVSV 523
QY 175 AAAADIISDVADIKGTLGEVWEFITHALNGKELWDLKLTGWVTGLFSRGWSNLESPFA 234
Db 524 AGAVIQVGSIGGALIGALANLTIIISNVIK-----KVSEWVSS-FSSGAQQAIAAKAA 575
QY 235 GVPGL-----TCATSGLSQVTLFGAAGLSASSGSLAHADSLASSASLPALAGTGG 284
Db 576 ELPGMQLSALANLMAIGLQAGKDLVOGLINGIGMVSANVNAKELASSVA-----GA 628
QY 285 GSGFGGLPSLAQVHAASLTQALRPADPGVGAABAEQVGGSQLVSAAGSQMGPGVGG 344
Db 629 VKGFLIGIBSPKLFTEYQFTAEFGNG-MEAGFKPIERAKDLAAELSRAM-----ESG 682
QY 345 MHPSSGASKGTTTKYSEGAAGTDEAERAPVE 377
Db 683 TDP-SGILAGLDQNELKQMLAALEERKRLKVE 714
RESULT 7
F70824
hypothetical glycine-rich protein Rv0747 - Mycobacterium tuberculosis (strain H37RV)
C: Species: Mycobacterium tuberculosis
C: Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C: Accession: F70824
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID: 98295987; PMID: 9634230
A: Accession: F70824
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-801 <COL>
A: Cross-references: GB:AL021958; GB:AL123456; NID: g3261536; PIDN: CAA17514.1; PID: g291102
A: Experimental source: strain H37RV
C: Genetics:
A: Gene: Rv0747
C: Superfamily: elastin
Query Match 7.7%; Score 152; DB 2; Length 801;

Best Local Similarity 28.1%; Pred. No. 0.034;
Matches 72; Conservative 23; Mismatches 111; Indels 50; Gaps 10;

QY 128 ALSAAFOAPFCAGAMAVVGALAYLVVVKTLINATQLLKLAKEAEIVAAAADIISDVA 186
DB : : : : :
67 AQAAAAFAOQF-VQAASGAGAYA-----AEEAAVSPLLAPINAQFVAATGRPLNGGA 119
:
QY 187 DIIKGTGLGEVWFITNALNKGELNKDKLTGWVTGLFSRGWSNLSEFPAGVPGLTGATSGL 246
DB : : : : :
120 NGAPGT-----GANG-----GPQGWLINGGGAGGS-----GAPGAGAGNG- 155
:
QY 247 SQVTGLFGAAGLSASSSLAHADSLAGSASLPAIAGIGGSGFGGLPSLAOVHAASTRQAL 306
DB -----GAGGLPSSGG-----AGCASTDVAGGAGGAGAG--NAGMLFGAAGVGV 199

QY 307 RPRADPGVGAABAFQVGGQSOLVSAQQSQMGVPPVMGMHPSSGASKGITTKKYSBEGAAA 366
DB : : : : :
200 GFSPNSGATCGAGGAGCAGGLFCAGRERGGSGGNLTGGAGGACGNAGTILAT--GDGGAG 257

QY 367 GTEDAERPVEADAGG 382
DB 258 CTCGARSRSGFFGGAGG 273

RESULT 8
G70917
hypothetical glycine-rich protein Rvi452c - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 07-Mar-2003
C:Accession: G70917
R:Cole, S.T.; Brosch, P.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.A.
RaJandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; PMID:98295987; EMBID:9634230
A:Accession: G70917
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
Molecule type: DNA
A:Molecular weight: 741 <COL>
A:Residues: GB:295844; GB:AL123456; NID:g3250713; PIDN:CAB09269.1; PID:g2131044
A:Cross-references: GB:295844; GB:AL123456; NID:g3250713; PIDN:CAB09269.1; PID:g2131044
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: Rvi452c
C:Superfamily: elastin related uncharacterized glycine-rich protein, PE motif containing

Query Match 7.7% Score 151; DB 2; Length 741;
Best Local Similarity 26.6%; Pred. No. 0.035;
Matches 80; Conservative 25; Mismatches 132; Indels 64; Gaps 13;

QY 110 EFVRPVADVDTYL-PVVGHALSAAFQPFPCAGAMAVV---GGALAYLVVKTLIN----- 159
DB : : : : :
9 ETVAASAASVDARICSSSIGVANSA-----AGSTTSVLAAADEVSAIAITLPESHAREY 62

QY 160 ---ATQLLKLLAKLAIELVAAAIIDI SDVADIINKTGLGEVWEFITNALN-----GLKE 209
DB : : : : :
63 QAISTOVAAPHDRFAQOTLSAAVSYVSABA-TNAAPTLEHNVLINALNAPTQALLCRPL 121

QY 210 LWDKL-----TGWVTGLFSRGWSNLSEFPAGVPGLTGATSGLSQTVCFLGAAGLSASSGL 264
DB : : : : :
122 IGDAAGAAGPTQAGGAGGILLWGNGGAGGSGAPGVGGAG----AADLFRTGGAGGAGGA 178

QY 265 AHADSCLASSASLPALAGIGGSGFGGLPSLAOVHAASTRQALRPADPDVGAAREQVGGQ 324
DB 179 GAAGGAGSGGW--LLGNGGVGGAGG-----QSL-----LCGATGGAGCN 216

QY 325 SOLVSAQSGSCQMGGPMVCMGMHPSSGAS--KGITTKKYSBGAACTEDAEARPVEDAGG 382
DB : : : : :
217 AGLFVGVTGTCGPCGPGCPGCVGTGGAGGLGGTYLAGAGHGAGGP-----GPIGGVGGH 271
YY 383 G 383 YY


```

QY 258 LSAAGSLAHADSLASSASLPALAGTGGSGGGGLPSLAQVHAASTROALRPAPGPGVAAA 317
Db 167 TGGSGGAGAAGGAGGSGGW--LYNGNGNGGGIGG-----NAIVAGGAGGNGGA 211
QY 318 ARQVGGQSOLVSAQSGSQGMPGVGMGHPSSGASKGTTTKKYSEGAAA----- 366
Db 212 ----GGAAGLWGGSGSGGGG----NGTGDGVNPAVPTNPALNGAAGDSNIEPQTSVL 263
QY 367 -GTEDAERAPVEADAGGG 383
Db 264 IGTGGGDTGGGAGVNGG 281

RESULT 15
D70835
hypothetical glycine-rich protein RV0278c - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: D70835
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.;
; Connor, R.; Davies, R.; Devlin, K.; Felkwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: D70835
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-957 <COL>
A:Cross-references: GB:AL021930; GB:AL123456; NID:93261524; PIDN:CAA17353.1; PID:g290946;
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: RV0278c
C:Superfamily: elastin

Query Match 7.5%; Score 148; DB 2; Length 957;
Best Local Similarity 27.9%; Pred.No. 0.074;
Matches 75; Conservative 13; Mismatches 108; Indels 73; Gaps 10;

QY 128 ALSAAPQAPFCAGAMVGGALAYLVWKTLLNATQLKLLAKLAELVAAAIADISVDAD 187
Db 64 ALSAQAQAFHAQFVQALISGGAY-----AAAERAAVSPLLDPINE 104

QY 188 IKTGLGEVBEFITHALNGKLKELDKLT-----CWVTGLFSRGWSNLESEFFAGVPGLT 240
Db 105 PFLANTGR--PLIIGANGAPG-----TCANGGDGWLIGNGGAGGS-----GAAGVN 150

QY 241 GATSGLSQVTGLFGAAGISASSGLAHADSLASSASLPALAGTGGSGGGGLPSLAQVHA 300
Db 151 GGAGNGGAGGNGGAGGLIGNGAGGAGGVASS-----GIGSGGGAGGNAMLFGAGGA 203

QY 301 STR----QALRPAPGPVCA-----AAROVGQ-----SQLVSAQSGGCGGGVGM 342
Db 204 GGAGGGVWALTGTGAGGAGGAGGAGNAGLIFGAAGVGGAGGFTNGSALGGAGGAGGAGGLFAT 263

QY 343 GGVHPSGASKGTTTKKYSEGAAAGTEDA 371
Db 264 GGVGSGGAG-----SSGAGGAGGA 284

Search completed: November 5, 2003, 17:00:58
Job time : 31 secs

```

Query Match 7.5%; Score 148; DB 2; Length 409;
 Best Local Similarity 24.2%; Pred. No. 0.028;
 Matches 94; Conservative 36; Mismatches 137; Indels 122; Gaps 15;

QY 12 SAIDGLYDLIGTIPNQGGI-----LYSSLEYFEKALEELAAAPGDCWLGSAAADKYAGNKR 68
 DB 29 AAWDGGAAEELGTAAQSFASVTAGLAGQAWGQPPAALAAAAAAYAGWLTAQAQAGSAG--- 85
 QY 69 NHVNFQQLADLDRLISLIHDQANAVQTRDILEGAKKGLFEVRPVAVDLTIPVVGHA 128
 DB 86 -----AACGARAVAS--IFEAQA-----TVLP----- 107
 QY 129 LSAAFQAPFCAGAMAVVGGALAYLVVKTLLINATQLLKLAKLAELVAAAIAIDISVDADI 188
 DB 108 -----AAVAANRDAFVLV-----MTNLFQGNAPLIAA----- 136
 QY 189 IXTGLGEVNEFITNALNGLKELMDKLTGVTGLFSRCWNSLSEFFAGV--PGLTGATSGLS 247
 DB 137 -SGVIEEMAADVAANSYYSGASATAQV-----PWASLLQRFPLCAGATGATGES 190
 QY 248 QVTGL-----FCAGLSAS-SGLAHADSLASSASI-----PALAGIGGSGRG 289
 DB 191 VGTGATGSGSVGTGGESVGTGATAGGGGVGVGVSAGLAGAAGPAHGVSQGVGNFG 250
 QY 290 GLPSLAQVHAASFEQALRPADGPPVCAAAEQVGGQSOLVSAQSGSQMGGPFVWGGMHPSS 349
 DB 251 GVGAGDVVASSATSA-----HAGVSPGFIGAPLALAAALGQVARGG-----T 293
 QY 350 GASKGTTTKKYSAGAAGTDEARAPVEA 378
 DB 294 NSAPGTATE-----SARAPEPAASAPPEA 317

RESULT 14
 F70971
 hypothetical glycine-rich protein RV3367 - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 31-Mar-2003
 C:Accession: F70971
 R:Cole, S.T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 395, 537-544, 1998
 A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Bartell, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A:Reference number: A70500; MUID:98295987; PMID:9634230
 A:Accession: F70971
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-538 <COL>
 A:Cross-references: GB:AL009198; GB:AL123456; NID:G3242262; PIDN:CAA15752.1; PID:el20228
 A:Experimental source: strain H37RV
 C:Genetics:
 A:Gene: RV3367
 C:Superfamily: uncharacterized glycine-rich protein, PE motif containing

Query Match 7.5%; Score 148; DB 2; Length 588;
 Best Local Similarity 24.5%; Pred. No. 0.042;
 Matches 78; Conservative 29; Mismatches 121; Indels 90; Gaps 13;

QY 119 LTVIPVVGHALSAFOAPFCAGAMVVGALAYLVVKTLLINATQLLKLAKLAELVAAAIA 178
 DB 1 MSFVVAVPEALAAA-----ASDVANIGSALSAANAAAAAGTT---GLLAAGADEVSAAL 51
 QY 179 ADIISDVA-----DIIKGTGL-----EVMFEITNALN----- 205
 DB 52 ASLFSGHAVSYQVAAQATALHDQFVALTGAGGSVALTEANVQQLNALINAPTQALL 111
 QY 206 GLKELMDKLTGWTG-----LFRGSWNSLSEFFAGVPLGTATSGLSQVTFGLFGAAG 257
 DB 112 GRPLIGDGAVGATSSPDGODGLLFFNGGAGYNS--AATPGMAGNGGNA---GLTNGG 166

Result No.	Score	Query			ID	Description
		Match	Length	DB		
1	153	7.8	836	1	VG26_BPM15	Q05233 mycobacteri
2	152	7.7	801	1	Y747 MYCTU	Q53810 mycobacteri
3	148.5	7.5	515	1	Y140 MYCTU	Q50594 mycobacteri
4	148	7.5	957	1	Y278 MYCTU	P56877 mycobacteri
5	145	7.4	914	1	WA22 MYCTU	Q066794 mycobacteri
6	142	7.2	778	1	Y324 MYCTU	P71933 mycobacteri
7	141.5	7.2	435	1	YU21 MYCTU	Q53268 mycobacteri
8	141	7.2	836	1	VG26_BPM2	P64220 mycobacteri
9	139.5	7.1	543	1	Y991 MYCTU	Q50630 mycobacteri
10	137.5	7.0	1901	1	Y208 MYCTU	Q53553 mycobacteri
11	132.5	6.7	463	1	Y468 MYCTU	Q53416 mycobacteri
12	132.5	6.7	864	1	ELS_RAT	Q09372 rattus norv
13	132.5	6.7	2333	1	PGCA_CANFA	P28343 canis fami
14	129.5	6.6	730	1	ELS_HUMAN	P15502 homo sapien
15	128.5	6.5	558	1	YJ83 MYCTU	Q10873 mycobacteri
16	127.5	6.5	1156	1	GLH4 CAEEL	Q78743 caenorhabdi
17	127	6.4	438	1	Y116 MYCTU	Q50615 mycobacteri
18	127	6.4	860	1	ELS_MOUSE	P54320 mus musculu
19	126.5	6.4	2038	1	FSH_DROME	P13709 drosophila
20	125.5	6.4	1758	1	CA24 CAEEL	P17140 caenorhabdi
21	125	6.3	603	1	YD25 MYCTU	Q10637 mycobacteri
22	124.5	6.3	1763	1	CA24_ASCSU	P27393 ascaris suu
23	124	6.3	1356	1	CA21_ORGNY	Q093484 oncorhynch
24	124	6.3	5263	1	PBOH_BOMMO	P05790 bombyx mori
25	123.5	6.3	747	1	SPD1_NEPFCL	P19937 nephtila cla
26	123	6.2	306	1	EC40 DAUCA	Q07322 daucus caro
27	122.5	6.2	1372	1	CA21_MOUSE	Q01149 mus musculu
28	121.5	6.2	1691	1	CA64 HUMAN	Q14031 homo sapien
29	120.5	6.1	1372	1	CA21_RAT	P02466 rattus norv
30	119.5	6.1	450	1	SWP1_ENGCCU	P092761 encephalito
31	119	6.0	864	1	KLTK_HUMAN	P29376 homo sapien
32	119	6.0	1707	1	CA24_MOUSE	P08122 mus musculu
33	119	6.0	2124	1	PGCA_RAT	P07897 rattus norv

Query Match	7.8%;	Score 153;	DB 1;	Length 836;
Best Local Similarity	22.3%;	Pred. No. 0.017;		
Matches 101;	Conservative 59;	Mismatches 157;	Indels 136;	Gaps 18;
Qy	13	AIDGLYDLIGIPNOGGIILSYSLVEYFEKALEELAAAPCGDGLGSAAADKYAGKRNHN	72	
Db	309	AMOGLSGTLG-----SVINLFNRUMESGLQAM-----GQLG-----GPLSTFIN	347	
Qy	73	FFOE-----LADLDROLISLIHDQANAVOTTRDILEGAKKG-LEFV	112	
Db	348	GFGLDFVSLMPALTSVSGLLIGNVLGTGLTQLAPIVTALTEAFQTLASTLTGMLTGALQAL	407	
Qy	113	RPVAVDLITYI-PVVGHALSAAPQ-----PPCAGAMAV	144	
Db	408	GPI---LTQVATLIGTLNTALQALQPMFLPSLMQCSFOQISDLVITSLAPHIPALATALGQ	464	
Qy	145	VGGA---LAVLVKTTLINATQLAKLAKLAEL-----V	174	
Db	465	VAGAVIQLADPTIIISTLTPA--FWQLVPKVAELVPTIVNLVQSFANLMPVVLPLAQALVSU	522	
Qy	175	AAAIADIISDVADIIKGTLGEVWEFTINALNGIKELWMDKLTGWVTGFLFSGWNSLEFFA	234	

[illegible][illegible]

16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical PE-PGRS family protein Rv0278c precursor.
GN Rv0278C OR MT0291 OR MTV035.06C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
CC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
CX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Colle S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gencies S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagals K., Krogh A., McLean J., Moule S., Murphy L.,
RA Hornsby S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RX FLEISCHMANN R.D., ALLAND C., EISEN J.A., CARPENTER L., WHITE O.,
RA PETERSON J., DEBOY R., DODSON R., GWINN M.L., HAFT D., HICKEY E.,
RA KELONAY J.F., NELSON W.C., UNAYAN L.A., EMOIAEVA M.D., SALZBERG S.L.,
RA DELICHER A., UTTERBACK T., WEIDMAN J., KHOURI H., GILL J., MIKULA A.,
RA BISHAI W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
CC SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

CC EMBL; AL021930; CAA17353.1; -
CC EMBL; AE006936; AAK4511.1; ALT_INIT.
CC PIR; D70835; D70835.
CC TIGR; MT0291, -
CC TubercuList; Rv0278c; -
CC InterPro; IPR000084; PE_region.
CC Pfam; PF00934; PE; 1.
CC Hypothetical protein; Repeat; Signal; Complete proteome.
FT SIGNAL 1 30
FT CHAIN 31 957
FT POTENTIAL.
FT HYPOTHETICAL PE-PGRS FAMILY PROTEIN
FT Rv0278C.
FT
FT CONFLICT 40 40 M -> I (IN REF. 2).
FT CONFLICT 158 163 MISSING (IN REF. 2).
FT CONFLICT 807 807 R -> G (IN REF. 2).
SQ SEQUENCE 957 AA; 81905 MW; 71EBABD417FEA47C CRC64;

Query Match 7.5%; Score 148; DB 1; Length 957;
Best Local Similarity 27.9%; Pred. No. 0.04;
Matches 75; Conservative 13; Mismatches 108; Indels 73; Gaps 10;

QY 128 ALSAQAQAPFCAGAMAVGGALAYLVVVKLTINATQLLKLLAKLAELVAAAIADIISVDAD 187
DB 64 ALSAQAQAPHAQFVQALTSGGAY-----AAAAAASVPELLOFINE 104
QY 188 IIKGTIGEWWEFTNALNGKLEWDKLT-----GWVTGLFGRGNSNLESFAGVFGLT 240
DB 105 FFLANTGR--PLIINGANGAPG---TGANGDGGWLITGNGAGGS-----GAAGVN 150
QY 241 GATSGHLSOVNGLFCAAGLSASSGLAHADSLASSASLPALAGICGGGSGFGGLPSLAOVHAA 300


```

GN Mycobacteriophage D29.
OS Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
OC NCBI_TaxID=28369;
ON [1]
RX SEQUENCE FROM N.A.
RP MEDLINE=98300335; PubMed=9636706;
RA Ford M.E., Sarkis G.J., Belanger A.E., Hendrix R.W., Hatfull G.F.;
RT "Genome structure of mycobacteriophage D29: implications for phage
RT evolution.";
RL J. Mol. Biol. 279:143-164 (1998).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF022214; AAC18467.1; --
CC DR PIR; H72802; H72802.
CC FT INIT MET 0 0 BY SIMILARITY.
CC SQ SEQUENCE 836 AA; 86572 MW; 528D12ED0FCB92E6 CRC64;
Query Match 7.2%; Score 141; DB 1; Length 836;
Best Local Similarity 22.3%; Pred. No. 0.095;
Matches 83; Conservative 57; Mismatches 137; Indels 96; Gaps 13;
QY 77 LADLRLQLSLHDOANAVQTRD-----ILEGAKKLEFVRPVAVDLTYIPVVGHALSAA 132
Db 370 LGTLGTQLAPIITLPAFTLADLTGLTALQALGVLTVVAE-----TLGTLALTA 424
QY 133 FQA-----PFCAGAMAVVGA---LAYLVKTLINA-- 160
Db 425 LQAIQMLPLVDSFKQLSETLVTSLGPLYPOIGFAFGVIGAVIQLAPTIISLIPAFQ 484
QY 161 -----TQLLKLLAKLAEV---AAAIADIISDVADIKIIGTLGEVW 197
Db 485 TLIPATAQLAPSLVQIVQFTKLPVIVPVGQVIVNLAADVQVQAGASTFLIGGISLV 544
QY 198 EFTNALNGKLKLDKLTGWTGLFSGWNSLESFFAGVPGLTGA-----TSGLS 247
Db 545 GVILADCVGVAE-----WV-GSWSSGVQVQVSDVFGQLPGKIKSWFDDAGSNLIEAGN 596
QY 248 QVTGLFGAAGLSASSGLAHADSLASASLIPALAGIGGGSGFGLPLSLAQVHAASRQALR 307
Db 597 VVQGLINGIGSMISSAVSRKELASSVK-----NAVTFGLGIHSPSRVFAEIGQTA 649
QY 308 PRADGPGVAAAEQVGQSQQLVSAQSGQMGPGVGMGHPSSGSGSKGTTTKYSEGAAG 367
Db 650 GFGNG-FEGFGQVIEKALAEUSQAM-----ESGVDP-SGILAGISTKELKYSAAL 702
QY 368 TDAERAPVEADA 380
Db 703 EQERKRIQVEKNA 715
RESULT 9
YP91_MYCTU STANDARD; PRT; 543 AA.
AC Q50630;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical PS-PORS family protein RV2591.
GN RV2591 OR MT2668.1 OR MTCY227.10C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Broach R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Das S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jags K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J.F., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PORS
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z77724; CAB01283.1; --
CC DR EMBL; AE007100; AAK46982.1; ALT_INIT.
CC DR PIR; F70726; F70726.
CC DR TIGR; MT2668.1; --
CC DR TubercuList; RV2591; --
CC DR InterPro; IPR000084; PE_region.
CC DR Pfam; PF00934; PE; 1.
CC DR ProDom; PD001223; PE_region; 1.
CC DR Hypothetical protein; Complete proteome.
KW SEQUENCE 543 AA; 46287 MW; 59730393E5D2DF59 CRC64;
SQ
Query Match 7.1%; Score 139.5; DB 1; Length 543;
Best Local Similarity 24.1%; Pred. No. 0.074;
Matches 77; Conservative 34; Mismatches 149; Indels 59; Gaps 11;
QY 94 AVOTTRDILGAKKLEFVRPVAVDLTYIPVVGHALSAAAFAPFCAGAMAVVGLAYLV 153
Db 33 AASTTSVLAAGAEVQAIARLFSVDVATHVQSLNQAALAHHSF-VQTLNAGG--AYSS 89
QY 154 VKTLINATQLLK--LILAKLAEVAAAADIISVDADIIRKTLGEVWEFTNALNGKELW 211
Db 90 AEAANASQAQLEQNLLA-----VINAPAQLFGR-----FLINGANGTAASP 132
QY 212 DKLTGWVTGLFSGWNSLESFFAGVPGLTGATSLQVTLFGAAGLSASSGLAHADSL- 270
Db 133 NGDGGI--LYNGNGNGFSGTGTAGVAGGAGGSLINGNGNGGAGGAGAGAGGAGGWL 190
QY 271 -----ASSASLPA-----LAGIGGSGFGLPLSL-----AQVHAASRQ 304
Db 191 LGNGGAGGPGGPTDVPAGTGTAGGAGGAGDAPLICGNGNGGPGGFAAFNGGAGGAGGAGS 250
QY 305 ALPRADGPFVGAABQVGGQSOLVSAQSGQMGPGVGMGHPSSGSGSKGTTTKYSEGA 364
Db 251 LFGVGGAGGVGGSSSDVGG---TGGAGGAGRGLFLGLGG---DGGAGGTNNNGDGG 302
QY 365 AAGTDAERAPVEADAGG 383
Db 303 AGGTAGRLFSLGDDGNG 321

```

```

RESULT 10
YZ08 MYCTU
ID YZ08 MYCTU STANDARD; PRT; 1901 AA.
AC O5353;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical PE-PGRS family protein RV3508 precursor.
GN RV3508 OR MTV023.15.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL022022; CAA17745.1;
CC FIR; F70806; F70806.
CC TubercuList; RV3508;
CC InterPro; IPR000084; PE_region.
CC Pfam; PF00934; PE; 1.
CC ProDom; PD001223; PE_region; 1.
CC Hypothetical protein; Repeat; Signal; Complete proteome.
FT SIGNAL 1 30
FT CHAIN 31 1901
FT STRAIN RV3508.
FT
FT SEQUENCE 1901 AA; 147627 MW; C7B1923D5D0146CD CRC64;
Query Match 7.0%; Score 137.5; DB 1; Length 1901;
Best Local Similarity 26.0%; Pred. No. 0.39;
Matches 56; Conservative 16; Mismatches 96; Indels 47; Gaps 5;
QY 212 DKLTGVTGLFGRGWSNLESFAGVPGTLTGATSGLSQVTLGFLCAAGLSASSGLAHADSL- 270
Db 341 DGVLGVGVTGKGK---VGVAGLGGAGGAGQLFSAGGAGAGVGGTGGGGGAGGAG 397
QY 271 ASASLAPALAGTGGSGF-----GGPLPSLAQVHAAS 302
Db 398 AAGADAPASTGLTGTGTFGAGGAGVGGGQGNIAAGGSGGAGGTGGGGAGGAGG 457
QY 303 ROALRPDAPGVG-----AAAEQVGGSQQLVSAQSGQGGPVGGMGHPSSGA 351
Db 458 DNASGIGADGGAGTGGNAGAGAGGAGGAGTGTGGVGAAGKAGTGGTGGGQGG-----AGG 513
QY 352 SKGTTTKYSEGAAGCTDAERAPVEADAGGCKY 386
Db 514 AGSAGTDAATGATGCTGTFSGGAGGAGGAGGNTGV 548

```

```

RESULT 11
YA68 MYCTU
ID YA68 MYCTU STANDARD; PRT; 463 AA.
AC O53416;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical PE-PGRS family protein RV1068c.
GN RV1068C OR MT1097 OR MTV017.21C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL021897; CAA17184.1;
CC EMBL; AE006991; AAK45353.1; ALT_INIT.
CC FIR; B70893; B70893.
CC TIGR; MT1097;
CC TubercuList; RV1068c;
CC InterPro; IPR000084; PE_region.
CC Pfam; PF00934; PE; 1.
CC ProDom; PD001223; PE_region; 1.
CC Hypothetical protein; Complete proteome.
FT CONFLICT 218 218
FT CONFLICT 235 235
FT G -> GGGAGIGAGDTGGDAGAGGAGGAGGAGGAGGAGG (IN REF. 2).
FT GGGGTGGGGGQGGVGGPDTGAAGGAGG (IN REF. 2).
FT
FT SEQUENCE 463 AA; 39305 MW; CF5696A7E9593952 CRC64;
Query Match 6.7%; Score 132.5; DB 1; Length 463;
Best Local Similarity 23.5%; Pred. No. 0.17;
Matches 72; Conservative 31; Mismatches 117; Indels 87; Gaps 10;
QY 119 LTYIPVVGHALSAFAFPFCAGAMAVVGAGLAVLWVKTLINATCLLKLAKLAEVAAI 178
Db 1 MSYMIAPDMLSSA-----AGDLASIGSSIN---ASTRAAAATATLLPAADEVSARI 51

```

CC 179 ADIIS-----DVADIIGTGLGEWFEITNAL 204
 CC 52 AALFSGHGEYQAIARQMAFHDQFTLALTSSAGAYASAEATNVQVGLI-NAPTOAL 110
 CC 205 NGKLEWDLKLTGWTC-----LPSRGSWLNLEFFAGVPGLTGATSGLSOVLTFQAA 256
 CC 111 LORPLTNGAGDTAANPNPAGAGGLLYGNGGFSQTTAGLTGCTGCGSAGLTGNGGAG 170
 CC 257 GLSASGLAHADSLASSASLALAGIGGSGFGGLPSLAQVHAASRQALRPADGPVGA 316
 CC 171 GAGANG-----AGNGGHWLYSGNGGAGGAGPAGAICA-----PGVAGGAGG 214
 CC 317 AABQVGGQSVLSAQSGQSGPVGNGMHPSSGASKGTTTKYKSEGAAGTDAERAPV 376
 CC 215 A-----GTTAGLFGNGGAGGAGGAGG-----RGGDGSAGWLSNGGDAGT----- 258
 CC 377 EADAGG 383
 CC 259 ----GGG 262

RESULT 12
 ELS_RAT
 ID ELS_RAT STANDARD; PRT; 864 AA.
 AC Q99372; 864 (Rel. 34, Created)
 DT 01-OCT-1995 (Rel. 34, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Elastin precursor (Tropoelastin) (Fragment).
 GN ELN.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Sutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RX MEDLINE=91104866; PubMed=1702999;
 RA Pierce R.A., Deak S.B., Stollie C.A., Boyd C.D.;
 RT "Heterogeneity of rat tropoelastin mRNA revealed by cDNA cloning";
 RL Biochemistry 29:9677-9683(1990).
 RN [2]
 RP SEQUENCE OF 781-864 FROM N.A.
 RX MEDLINE=88330868; PubMed=2971041;
 RA Deak S.B., Pierce R.A., Belsky S.A., Riley D.J., Boyd C.D.;
 RT "Rat tropoelastin is synthesized from a 3.5-kilobase mRNA";
 RL J. Biol. Chem. 263:13504-13507(1988).
 RN [3]
 RP SEQUENCE OF 264-533 AND 558-864 FROM N.A., AND ALTERNATIVE SPLICING.
 RX MEDLINE=92241859; PubMed=1572637;
 RA Pierce R.A., Alatawi A., Deak S.B., Boyd C.D.;
 RT "Elements of the rat tropoelastin gene associated with alternative splicing";
 RL Genomics 12:651-658(1992).
 CC -!- FUNCTION: MAJOR STRUCTURAL PROTEIN OF TISSUES SUCH AS AORTA AND
 CC NUCAL LIGAMENT, WHICH MUST EXPAND RAPIDLY AND RECOVER COMPLETELY.
 CC -!- SUBUNIT: THE POLYMERIC ELASTIN CHAINS ARE CROSS-LINKED TOGETHER
 CC INTO AN EXTENSIBLE 3D NETWORK.
 CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX OF ELASTIC FIBERS.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=alternative splicing; Named isoforms=8;
 CC Comment=Experimental confirmation may be lacking for some
 CC isoforms;
 CC Name=1;
 CC IsoId=Q99372-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q99372-2; Sequence=VSP_004244;
 CC Name=3;
 CC IsoId=Q99372-3; Sequence=VSP_004245;
 CC Name=4;
 CC IsoId=Q99372-4; Sequence=VSP_004246;
 CC Name=5;
 CC IsoId=Q99372-5; Sequence=VSP_004244, VSP_004245;

CC Name=6;
 CC IsoId=Q99372-6; Sequence=VSP_004245, VSP_004246;
 CC Name=7;
 CC IsoId=Q99372-7; Sequence=VSP_004244, VSP_004246;
 CC Name=8;
 CC IsoId=Q99372-8; Sequence=VSP_004244, VSP_004245, VSP_004246;
 CC -!- PTM: THE CROSSLINKS ARE MADE OF DEAMINATED LYS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M60647; AAA42269.1; -;
 CC EMBL; J04035; AAA42268.1; -;
 CC EMBL; M8372; AAA42271.1; -;
 CC EMBL; M8355; AAA42271.1; JOINED.
 CC EMBL; M8363; AAA42271.1; JOINED.
 CC EMBL; M8364; AAA42271.1; JOINED.
 CC EMBL; M8366; AAA42271.1; JOINED.
 CC EMBL; M8371; AAA42271.1; JOINED.
 CC EMBL; M8376; AAA42272.1; -;
 CC EMBL; M8373; AAA42272.1; JOINED.
 CC EMBL; M8375; AAA42272.1; JOINED.
 CC PIR; A36106; EART.
 CC InterPro; IPR003979; tropoelastin.
 CC PRINTS; PR01500; TROPOELASTIN.
 CC Structural Protein; Connective tissue; Repeat; Signal;
 KW Alternative splicing.
 FT NON TER 1 1
 FT SIGNAL <1 21 BY SIMILARITY.
 FT CHAIN 22 864 ELASTIN.
 FT DISULFID 854 859 BY SIMILARITY.
 FT VARSPPLIC 263 307 Missing (in isoform 2, isoform 5, isoform
 7 and isoform 8).
 FT VARSPPLIC 308 308 /FTID=VSP_004244.
 FT VARSPPLIC 809 823 Missing (in isoform 3, isoform 5, isoform
 6 and isoform 8).
 FT VARSPPLIC 809 823 /FTID=VSP_004245.
 FT VARSPPLIC 809 823 Missing (in isoform 4, isoform 6, isoform
 7 and isoform 8).
 FT VARSPPLIC 809 823 /FTID=VSP_004246.
 SQ SEQUENCE 864 AA; 72786 MW; 456894BB09E79FD4 CRC64;
 Query Match 6.7%; Score 132.5; DB 1; Length 864;
 Best Local Similarity 23.1%; Pred. NO. 0.34; Indels 143; Gaps 18;
 Matches 94; Conservative 35; Mismatches 135;
 Qy 9 PTISAIDGLYDLGLI-GIPNOGGILYSSLEYFEKALEELAAAPPGDGLG----- 57
 Db 384 PGVGGIPGVGGIPGVGGIPGVGG-----PGIGGPGIVGGPGAVSP 423
 Qy 58 SAADKTAGKRNHNFFQELADLDROLISLIHQANAVQTTDRDLEGAKKGLFVFPVAV 117
 Db 424 AAAAKAAAKAAY-----GARGVG----- 443
 Qy 118 DLATVVPVGHASAAFOAPFCAGAMAVVGGALAYLVVKTLINATQLLKLAELAAVAAA 177
 Db 444 ----IPTYG--VGAGGPPGVGVGAGAGLGA-----SQAAAA 475
 Qy 178 IADIISDVADIIGTGLGEWFEITNALNGKLWDLKLTGVTGLFSRGSWLNLEFFAGV 237
 Db 476 AA---AKAAYGAGGAGTGLGLVPGAVPG-----ALPGAVPGALP-----GAVP 516
 Qy 238 G-LTGATSGLSQVTLFGAAGLSASSGLAHADSLASSASL---PALAGIGGSGFGGLP- 292
 Db 517 GALPGAVPGVPGTGGVPGAGTTPAAAAAAXAAXAAXAAXAAXAAXAAXAAXAAXAAX 576
 Qy 293 --SLAQVHAASL--RQALRPADGPVGAABQVGGOSQLVSAQSGQ-----GMGGP-V 340

Db 577 GVPGGVGTGTGTGLVP--GDLGGAGTAAKSAKAAKAAQYRAAGLGAAGVPL 633
QY 341 GMGMHPSGA-----SKGTTTKYSEGAAGTDAEPVADAG 382
Db 634 GVGAGVFGAGAGGAGGAGVFCGAGAVPGSLAASKAKYGAAG 680

RESULT 13
PGCA_CANFA STANDARD; PRT; 2333 AA.
AC Q28343; Q28310;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Aggrecan core protein precursor (Cartilage-specific proteoglycan core
DE protein) (CSFPCP).
GN AGC1.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RN [2]
RN [3]
RN [4]
RN [5]
RN [6]
RN [7]
RN [8]
RN [9]
RN [10]
RN [11]
RN [12]
RN [13]
RN [14]
RN [15]
RN [16]
RN [17]
RN [18]
RN [19]
RN [20]
RN [21]
RN [22]
RN [23]
RN [24]
RN [25]
RN [26]
RN [27]
RN [28]
RN [29]
RN [30]
RN [31]
RN [32]
RN [33]
RN [34]
RN [35]
RN [36]
RN [37]
RN [38]
RN [39]
RN [40]
RN [41]
RN [42]
RN [43]
RN [44]
RN [45]
RN [46]
RN [47]
RN [48]
RN [49]
RN [50]
RN [51]
RN [52]
RN [53]
RN [54]
RN [55]
RN [56]
RN [57]
RN [58]
RN [59]
RN [60]
RN [61]
RN [62]
RN [63]
RN [64]
RN [65]
RN [66]
RN [67]
RN [68]
RN [69]
RN [70]
RN [71]
RN [72]
RN [73]
RN [74]
RN [75]
RN [76]
RN [77]
RN [78]
RN [79]
RN [80]
RN [81]
RN [82]
RN [83]
RN [84]
RN [85]
RN [86]
RN [87]
RN [88]
RN [89]
RN [90]
RN [91]
RN [92]
RN [93]
RN [94]
RN [95]
RN [96]
RN [97]
RN [98]
RN [99]
RN [100]

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL; U65989; AAB06238.2; -;
DR EMBL; S74662; AAC60527.1; -;
DR EMBL; L07054; -; NOT_ANNOTATED_CDS.
DR PIR; I46998; I46998.
DR HSP; P08709; I46998.
DR InterPro; IPR000152; Asx hydroxyl.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003006; Ig MHC.
DR InterPro; IPR000538; Lactin_C.
DR InterPro; IPR001304; Lactin_C.
DR InterPro; IPR003324; SGXSG.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00009; EGF; 1.
DR Pfam; PF00047; Ig; 1.
DR Pfam; PF00059; Lactin_C; 1.
DR Pfam; PF02339; SGXSG; 66.
DR Pfam; PF00084; sushi; 1.
DR Pfam; PF00193; Xlink; 4.
DR PRINTS; PR01365; LINKMODULE.
DR PRINTS; PR00356; ANTIFREEZEII.
DR PRODOM; PD000918; Link; 4.
DR SMART; SM00032; CCP; 1.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00179; EGF_CA; 1.
DR SMART; SM00445; LINK; 4.
DR PROSITE; PS00010; ASX HYDROXYL; 1.
DR PROSITE; PS00615; C-TYPE LECTIN_1; 1.
DR PROSITE; PS00041; C-TYPE LECTIN_2; 1.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS00835; IG LIKE; 1.
DR PROSITE; PS00290; IG MHC; 1.
DR PROSITE; PS01241; LINK; 4.
KW Glycoprotein; Cartilage; Proteoglycan; Lactin; Signal; Sushi;
KW EGF-like domain; Repeat; Immunoglobulin domain.
FT SIGNAL 1 16
FT CHAIN 17 2333 AGGREGAN CORE PROTEIN.
FT DOMAIN 34 147 IG-LIKE V-TYPE.
FT DOMAIN 170 247 LINK 1.
FT DOMAIN 268 349 LINK 2.
FT DOMAIN 513 590 LINK 3.
FT DOMAIN 611 692 LINK 4.
FT DOMAIN 2081 2117 EGF-LIKE, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 2130 2245 C-TYPE LECTIN.
FT DOMAIN 2249 2307 SUSHI.
FT DOMAIN 48 140 G1-A.
FT DOMAIN 152 247 G1-B.
FT DOMAIN 253 349 G1-B'.
FT DOMAIN 495 589 G2-B.
FT DOMAIN 596 691 G2-B'.
FT DOMAIN 694 816 KS.
FT DOMAIN 819 1394 CS-1.
FT DOMAIN 1395 2079 CS-2.
FT DOMAIN 2080 2333 G3.
FT DISULFID 51 133 BY SIMILARITY.
FT DISULFID 175 246 BY SIMILARITY.
FT DISULFID 199 220 BY SIMILARITY.
FT DISULFID 273 348 BY SIMILARITY.
FT DISULFID 297 318 BY SIMILARITY.
FT DISULFID 518 589 BY SIMILARITY.
FT DISULFID 542 563 BY SIMILARITY.
FT DISULFID 616 691 BY SIMILARITY.
FT DISULFID 640 661 BY SIMILARITY.
FT DISULFID 2085 2096 BY SIMILARITY.

```
FT DISULFID 2090 2105 BY SIMILARITY.
FT DISULFID 2107 2116 BY SIMILARITY.
FT DISULFID 2123 2134 BY SIMILARITY.
FT DISULFID 2151 2243 BY SIMILARITY.
FT DISULFID 2219 2235 BY SIMILARITY.
FT DISULFID 2250 2293 BY SIMILARITY.
FT DISULFID 2279 2306 BY SIMILARITY.
FT CARBOHYD 126 128 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 239 239 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 333 333 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 387 387 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 444 444 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 620 620 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 676 676 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 747 747 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 2333 AA; 240573 MW; 8B9ED78F3508B596 CRC64;

Query Match 6.7%; Score 132.5; DB 1; Length 2333;
Best Local Similarity 26.4%; Pred. No. 1;
Matches 69; Conservative 35; Mismatches 82; Indels 75; Gaps 13;

Qy 172 ELVAAAIADI-----ISDVADIKTLGLGVWFFITNALNGLKELWDLKLTGMVT 219
Db 940 EVASGVEDLSGLPSGEGPETSTSGVGDLSRLPSGEGPEVSASGV-----GDL 988
Qy 220 GLFSRGSNLESFAGVPLTG-----ATSLGSQVTL-----FGAAGLSASGG 263
Db 989 GLPS-GREGLETSTGVEDLSGLPSGEGPEASTSGVGDLSRLPSGEGPEVSASGVEDLSG 1047
Qy 264 LAHADSLASAS-LPALAGIGG-----SFGGLPSLAQVHAATRCALPRADGP-V 314
Db 1048 LPSGEGLEASAGVGLSLPSGEGPEASAGVGDLSRL-----PSGEGPEV 1094
Qy 315 GAAAEQ-----VGGQQLVSAQSQMGQ-PVGMGMHPSS---GASKGTTTKYSEGA 365
Db 1095 SASGVEDLSGLSGESPEASAGVGDLSGLPSGREGLETASAGVGDLSGLPSGEGQEA 1154
Qy 366 AGTDAERAP-----VEADAGG 382
Db 1155 SGVEDLSRLPSGEGPEASAG 1175

RESULT 14
ELS_HUMAN
ID ELS_HUMAN STANDARD; PRT; 730 AA.
AC P15502; Q14233; Q14238;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Elastin precursor (Tropoelastin).
GN ELN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM B).
RX MEDLINE=87289668; PubMed=3039501;
RA Indik Z., Yeh H., Ornstein-Goldstein N., Sheppard P., Anderson N.,
RA Rosenbloom J.C., Felton L., Rosenbloom J.;
RA Ornstein-Goldstein N., Yeh H., Rosenbloom J., Uitto J.;
RT "Alternative splicing of human elastin mRNA indicated by sequence
RT analysis of cloned genomic and complementary DNA.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:5680-5684(1987).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=8900960; PubMed=3171221;
RA Fazio M.J., Olsen D.R., Kauh E.A., Baldwin C.T., Indik Z.,
RA Ornstein-Goldstein N., Yeh H., Rosenbloom J., Uitto J.;
RT "Cloning of full-length elastin cDNAs from a human skin fibroblast
RT recombinant cDNA library: further elucidation of alternative splicing
RT utilizing exon-specific oligonucleotides.";
RL J. Invest. Dermatol. 91:458-464(1988).
```

```
RN RP SEQUENCE OF 164-724 FROM N.A. (ISOFORM B).
RC TISSUE=Placenta;
RX MEDLINE=88156136; PubMed=2831431;
RA Fazio M.J., Olsen D.R., Kuivaniemi H., Chu M.L., Davidson J.M.,
RA Rosenbloom J., Uitto J.;
RT "Isolation and characterization of human elastin cDNAs, and age-
RT associated variation in elastin gene expression in cultured skin
RT fibroblasts.";
RL Lab. Invest. 58:270-277(1988).
RN [4]
RP SEQUENCE OF 603-730 FROM N.A.
RC TISSUE=Hippocampus, and Placenta;
RX MEDLINE=96291399; PubMed=8689688;
RA Frangiskakis J.M., Ewart A.K., Morris C.A., Mervis C.B.,
RA Bertrand J., Robinson B.F., Klein B.P., Ensing G.J., Everett L.A.,
RA Green E.D., Proeschel C., Gutowski N.J., Noble M., Atkinson D.L.,
RA Odeberg S.J., Keating M.T.;
RT "LIM-kinase hemizyosity implicated in impaired visuospatial
RT constructive cognition.";
RL Cell 86:59-69(1996).
CC -!- FUNCTION: MAJOR STRUCTURAL PROTEIN OF TISSUES SUCH AS AORTA AND
CC NUCHAL LIGAMENT, WHICH MUST EXPAND RAPIDLY AND RECOVER COMPLETELY.
CC -!- SUBUNIT: THE POLYMERIC ELASTIN CHAINS ARE CROSS-LINKED TOGETHER
CC INTO AN EXTENSIBLE 3D NETWORK.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX OF ELASTIC FIBERS.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Comment=Additional isoforms seem to exist;
CC Name=1;
CC IsoId=P15502-1; Sequence=Displayed;
CC Name=2;
CC IsoId=P15502-2; Sequence=VSP_004243;
CC -!- PTM: THE CROSSLINKS ARE MADE OF DEAMINATED LYS.
CC -!- DISEASE: Haploinsufficiency of ELN may be the cause of certain
CC cardiovascular and musculo-skeletal abnormalities observed in
CC Williams-Beuren syndrome (WBS), a rare developmental disorder. It
CC is a contiguous gene deletion syndrome involving genes from
CC chromosome band 7q11.23.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```

DR MTM; 194050; --
DR GO; GO:0005578; C:extracellular matrix; TAS.
DR GO; GO:0005615; C:extracellular space; TAS.
DR GO; GO:0005201; F:extracellular matrix structural constituent; TAS.
DR GO; GO:0008283; P:cell proliferation; TAS.
DR GO; GO:0008015; P:circulation; TAS.
DR GO; GO:0007397; P:histogenesis and organogenesis; TAS.
DR GO; GO:0007585; P:respiratory gaseous exchange; TAS.
DR InterPro; IPR003979; tropoelastin.
DR PRINTS; PR01500; TROPOLASTIN.
KW Structural protein; Connective tissue; Repeat; Signal;
KW Williams-Beuren syndrome; Alternative splicing.
FT SIGNAL 1 26
FT CHAIN 27 730 ELASTIN.
FT DISULFID 720 725 BY SIMILARITY.
FT VARSPLIC 472 477 Missing (in isoform 2).
FT FTID=VSP_004243.
SQ SEQUENCE 730 AA; 63260 MW; AB06D155A567AE46 CRC64;

Query Match 6.6%; Score 129.5; DB 1; Length 730;
Best Local Similarity 20.3%; Pred. No. 0.44;
Matches 96; Conservative 36; Mismatches 134; Indels 167; Gaps 15;

QY 22 GIGPN--GGTLYSSLEFEKALEELAAFPDGLGSAADKYAGKRNHNVPQELAD 79
Db 143 GVLPGVYGGVL-----PGARFFGVGLPGV----- 169
QY 80 LDRQLISLHDOANAVQTTDILEGAKGLEFVRPV-----AVDLTYIPVVGHALSAAFO 134
Db 170 -----PTGAGVKPKAPGVGGAFAGIPGVGPGPGQVPLGY-PIKAPKLPGGYG 218
QY 135 AFGCAGM-----AVGGA-LAYLVVTLNATQLKLAKLAEVAAIAADIISDVA 186
Db 219 LFYTTKLPYGPVGGVAGAGAGYPTGTGVQAAAAAATAKAGAGAGVLPVGV 278
QY 187 DIIKGTGLGEVWFITNALNGLXELNDKLTGWITGLFSRGWSNLEFFAGVPLGTATSG 246
Db 279 G-----AGVGVPGA----- 288
QY 247 SQVTGLFGAAGLSASGLAHADSLASSSLPALAG-IGGSGSGF-----GLPSL-- 294
Db 289 --IPGIGGIAGVGTAAAAAATAKAAKAGVGAAGLVPGGPGFGVGVGVGGVGVGV 346
QY 295 -----AQVHAASTQALRPADGPVGAARQVG----- 322
Db 347 PGAGIPVPGAGIPGAAVPGVVSPEAAAAKAAKAGARPGVGVGGIPTVGVGAGGPF 406
QY 323 -----GQSOLVSAQSGQGVPGVGMGMPSSGASKGTTTKYSEG--AAAGTDA 371
Db 407 FGVGVGGIFGVAGVPGVGVGVGVGV-GISPEAQAAAAKAAKAGYGVGTAAAAAKAA 465
QY 372 ERA 374
Db 466 AKA 468

RESULT 15
YJ83_MYCTU
ID YJ83_MYCTU STANDARD; PRT; 558 AA.
AC Q10873;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical PE-PGRS family protein Rv1983.
GN Rv1983 OR NT2036 OR MTCv39.36C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;

```

```

RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala E.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies K., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean A., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sultun J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; Z74025; CAA98400.1; --
DR EMBL; AE007056; AAK46311.1; --
DR PIR; E70756; E70756.
DR TIGR; MT2036; --
DR TubercuList; Rv1983; --
DR InterPro; IPR000084; PE_region.
DR Pfam; PF00934; PE; 1.
DR ProDom; PD001223; PE_region; 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT DOMAIN 27 53 ALA-RICH.
FT TRANSMEM 332 352 POTENTIAL.
FT TRANSMEM 360 380 POTENTIAL.
SQ SEQUENCE 558 AA; 53738 MW; 17ECBE43778E021E CRC64;

Query Match 6.5%; Score 128.5; DB 1; Length 558;
Best Local Similarity 25.4%; Pred. No. 0.38;
Matches 71; Conservative 23; Mismatches 125; Indels 61; Gaps 8;

QY 110 EFVRPVADLTYPVVGHALSAAPQAFPCAGAVA---VVGALA-----YLVYKTL 157
Db 9 EFLTSAADVENIGSTLRANAATAAATAAAGADEVSAVAALFARFGQYCAVSAQ 68
QY 158 INA--TOLLKLAKLAEVAAIAADIISDVADIKLTGLGEVWFITNALNGLXELNDKLT 215
Db 69 ASAPHQQFVQTLNSASGSYAAAEATIASQQTQHQHLLGAV-NAPTETLLGRLLIGDAP 127
QY 216 GWTG-----LFSRGWSNLEFFAGVPLGTATSGLSQVTLFGAAGLSASSGLAHA 267
Db 128 GTATSPNGGAGGLLYGNGNGYGSATASGVGGAGGSAGLLGNGGAGGAGGPNAGGAG-- 185
QY 268 DSLASSASLPALAGIGGSGFGGLPSLAQVHAASTQALRPADGPVGAARQVGGSQ 327
Db 186 ---GNGWLLGNGGIGGPGGASSIPGMS-----GGAGGT 216
QY 328 VSAQSGCGMGPVGMGMHPSSGASKGTTTKYSEGAAAG 367
Db 217 GGAAGLLGWANGAGGLGDGVGVDRCT-----GGAGG 249

Search completed: November 5, 2003, 17:00:21

```

Thu Nov 6 11:33:28 2003

us-09-688-672a-14.rsp

Page 12

Job time : 23 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model
Run on: November 5, 2003, 17:01:02 ; Search time 56 Seconds
(without alignments)
1806.368 Million cell updates/sec

Title: US-09-688-672A-14
Perfect score: 1969
Sequence: 1 MSRAIIPTISAIDGLYDL.....RAPVEADAGGQKVLVRNV 392

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL_23.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rv.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1969	100.0	392	16	O6267 mycobacteri
2	1258	63.9	394	16	Q49722 mycobacteri
3	380.5	19.3	402	16	P96213 mycobacteri
4	168	8.5	496	16	O53631 mycobacteri
5	168	8.5	533	16	Q8VKR5 mycobacteri
6	156.5	7.9	536	16	Q8FND6 corynebacte
7	155	7.9	1329	16	O68810 mycobacteri
8	152	7.7	1408	16	Q8VKI17 mycobacteri
9	151	7.7	738	16	Q8VKI15 mycobacteri
10	151	7.7	741	16	O68808 mycobacteri
11	150	7.6	1306	16	O53775 mycobacteri
12	149.5	7.6	916	9	O80116 bacterioph
13	149	7.6	628	16	Q8VJI9 mycobacteri
14	148.5	7.5	749	16	O53844 mycobacteri
15	148	7.5	409	16	P95182 mycobacteri
16	148	7.5	588	16	O50396 mycobacteri

ALIGNMENTS

RESULT 1

ID O6267 PRELIMINARY; PRT; 392 AA.

AC O6267;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein Rv3616C
GN Rv3616C OR WT3718 OR MTCY07H7B.06.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

17	148	7.5	1665	16	O53215	mycobacteri
18	147.5	7.5	459	16	Q8XQX4	raistonia s
19	147	7.5	436	16	Q8VIX6	mycobacteri
20	147	7.5	439	16	O06287	mycobacteri
21	146.5	7.4	837	16	O53684	mycobacteri
22	145.5	7.4	1507	16	Q8VJ23	mycobacteri
23	145.5	7.4	1538	16	O53395	mycobacteri
24	145	7.4	911	16	Q8DXK6	streptococ
25	144.5	7.3	539	16	F71857	mycobacteri
26	144	7.3	1381	16	O53552	mycobacteri
27	144	7.3	1384	16	Q8VIZ1	mycobacteri
28	142.5	7.2	454	16	Q8VJD7	mycobacteri
29	142.5	7.2	525	16	O53284	mycobacteri
30	142.5	7.2	667	16	O53415	mycobacteri
31	142.5	7.2	879	16	Q8VKD2	mycobacteri
32	142.5	7.2	882	16	O53845	mycobacteri
33	142	7.2	476	16	O06199	mycobacteri
34	141.5	7.2	1489	16	O53559	mycobacteri
35	141	7.2	584	16	O53575	mycobacteri
36	141	7.2	1468	5	Q9GUB5	galleria me
37	140.5	7.1	853	16	O53439	mycobacteri
38	140	7.1	1463	16	Q9ADP6	streptomyc
39	138.5	7.0	614	10	Q9LH97	arabidopsis
40	138.5	7.0	635	16	Q8VK71	mycobacteri
41	138.5	7.0	683	16	Q8G791	bifidobacte
42	137.5	7.0	783	16	O53809	mycobacteri
43	137.5	7.0	1715	16	Q8VIZ0	mycobacteri
44	136	6.9	476	16	Q8VKA5	mycobacteri
45	136	6.9	618	16	O06801	mycobacteri

```

DR EMBL; Z95537; CAB08950.1; -.
DR ENBL; AB007171; AAK48077.1; -.
DR TIGR; MT3718; -.
DR TubercuList; RV3616c; -.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 192 192 T -> I (IN REF. 2).
SQ SEQUENCE 392 AA; 39888 MW; 82BBA8DD9D6F567 CRC64;

Query Match 100.0%; Score 1969; DB 16; Length 392;
Best Local Similarity 100.0%; Pred. No. 7.9e-119;
Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSRAFIIDPTISAIDGLYDLGIGIPNQGILYSSLEYFEKALEBELAAAFPGDGLGSA 60
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1 MSRAFIIDPTISAIDGLYDLGIGIPNQGILYSSLEYFEKALEBELAAAFPGDGLGSA 60

QY 61 DYAGKRNHNHVPFQELADRLQLISLIHQANAVQTTTRDILEGAKKGLFVRPVAVDLT 120
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 61 DYAGKRNHNHVPFQELADRLQLISLIHQANAVQTTTRDILEGAKKGLFVRPVAVDLT 120

QY 121 YIPVGHASAAFPQAFPCAGAMAVGGALAYLVKTLINATOLLKLAELVAAATAD 180
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 121 YIPVGHASAAFPQAFPCAGAMAVGGALAYLVKTLINATOLLKLAELVAAATAD 180

QY 181 IISDVADIITKTLGEVWEFTNALNGLKELWDKLTGWVTGLFSRGSNLESFFAGVPGLT 240
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 181 IISDVADIITKTLGEVWEFTNALNGLKELWDKLTGWVTGLFSRGSNLESFFAGVPGLT 240

QY 241 GATSGLSQVTLFGAAGLSASSGLAHADSLASSASLPALAGIGGGSGGGLPSLAQVHAA 300
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 241 GATSGLSQVTLFGAAGLSASSGLAHADSLASSASLPALAGIGGGSGGGLPSLAQVHAA 300

QY 301 STRQALRPRADGPVGAARAEQVGSQSLVSAQSQGCGPVGMGVHPSPSSGSKGTTTKY 360
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 301 STRQALRPRADGPVGAARAEQVGSQSLVSAQSQGCGPVGMGVHPSPSSGSKGTTTKY 360

QY 361 SEGAAAGTDAERAPVEADAGGQKVLVRNVV 392
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 361 SEGAAAGTDAERAPVEADAGGQKVLVRNVV 392

RESULT 2
Q49722 PRELIMINARY; PRT; 394 AA.
ID P96213 AC Q49722;
AC Q49722; (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DE Hypothetical protein MLO405.
GN MLO405 OR 31620 C2 213 OR MLC1383.01.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Horsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Ouail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus."
RL Nature 409:1007-1011(2001).
RN [2]
RP SEQUENCE FROM N.A.
RA Smith D.R., Robison K.;
RA Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.

```

RESULT 4	
O53631	PRELIMINARY; PRT; 496 AA.
ID	O53631
AC	O53631;
DT	01-JUN-1998 (TrEMBLrel. 06, Created)
DT	01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT	01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE	PGRS-family protein.
GN	RV0109 OR NTV031.03C.
OS	Mycobacterium tuberculosis.
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetaceae; Mycobacteriaceae; Mycobacterium

RESULT 5	
OSVKRS	PRELIMINARY; PRT; 533 AA.
ID	
AC	OSVKRS; 01-MAR-2002 (T-EMBLrel. 20, Created)
DT	01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT	01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE	PE PGKS family protein.
GN	MT0118.
OS	Mycobacterium tuberculosis.
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC	Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX	NCBI_TaxID=1773;
EN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=CDC 1551 / Oshkosh;
RA	F. Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.
RA	Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.
RA	Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg

```

RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AS006923; AAK4341.1; -.
DR TIGR; MT0118; -.
DR InterPro; IPR002952; Eggshell.
DR InterPro; IPR000084; PE_region.
DR Pfam; PF00934; PE; 1.
DR PRINTS; PR01228; EGGSHLL.
DR ProDom; PD001223; PE_region; 1.
DR ProDom; PD001223; PE_region; 1.
SQ SEQUENCE 533 AA; 46785 MW; 56F0115F01F4D4E5 CRC64;

Query Match      8.5%; Score 168; DB 16; Length 533;
Best Local Similarity 26.2%; Pred. NO. 0.009;
Matches 88; Conservative 24; Mismatches 140; Indels 84; Gaps 14;

QY 82 RLISLIHQANAVQTRDILEGAKKLEFVR-PVAVDLTYIPVVG-----126
Db 35 RSLMSLIITSPATAAATHLAGISALSTANAARAAAPTALSVADEVSVLIAALFEA 94
QY 127 -----HALSA---AFQAPFCAGAMVVGALAYLVVKTLLINATQLLKLAKLAELVAAAI 178
Db 95 YAQEYQALSAQALAFHDQF---VQALNMGAVCYAAAEI-ANATPLQAL-----QTVQONV 145
QY 179 ADIISDVADIILKTLGEVVEFTNALNGLKELWDLKLTGWVTG---LFSRGSNLSFFA 234
Db 146 LTVVNAPTQALLGR-----PIIGNGANLPN-----TGQDGGPGLLFCNGN-----188
QY 235 GVPGLTCAATGSLGSLQVTLFCAAGL---SASSGLAHADSLASSALPA-----LAGIGGGSG 287
Db 189 -----GGSGGVDAQGNGGAAGLGNCGSGGVGGVGGIAGSAGGAGGALLFGNGGPGG 242
QY 288 FGCLPSLAQVHAASRASTRQALRPADGPVGAASQVGGQSQQLVSAQSGQMGVPGVGGMHP 347
Db 243 AGGIGTGD-----GGPGGAGNAIG-----LFGSGTGGGNGVGGVGGVGN 284
QY 348 SSQASKGTTTKYSEGAAGTDEARPEADAGGG 383
Db 285 GGNAGNGGTAGLFGHGGAGGAGGIGSADGGLGGGG 320

RESULT 6
Q8FND6 PRELIMINARY; PRT; 536 AA.
AC Q8FND6
DT 01-MAR-2003 (TEMBLrel. 23, Created)
DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN CE2208.
OS Corynebacterium efficiens.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Corynebacteriidae; Corynebacterium.
OX NCBI_TaxID=152794;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;
RA Kawarabayasi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,
RA Ikeo K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,
RA Tsuda Y., Sugimoto S.;
RT "The entire genomic sequence of Corynebacterium efficiens YS-314.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF005221; BAC19018.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 536 AA; 52889 MW; A6231609CCCB2114 CRC64;

Query Match      7.9%; Score 156.5; DB 16; Length 536;
Best Local Similarity 23.3%; Pred. NO. 0.05;
Matches 99; Conservative 48; Mismatches 163; Indels 115; Gaps 20;

QY 42 ALLELAAPFG--DGLWLSAADKYAGKRNHNVFFQEL-ADLDRLSLIHQANAV--Q 96

```

```

Db 126 SLDOLCSDFSGTNSGAVSAAQSSWT---TMASTISEVSASLDRVAGELLASNAGEVFQ 181
QY 97 TTRDILEGAKKG-----LEFVRPVAVDLTYPVVGHALSAAFQAPFCAGAMAVVGGAL 149
Db 182 ASARIAEVAEAGSVFSONAREMSRSVGT-LNQI-YNGHKMS-VFMA---AASIAMIKEPA 235
QY 150 AYLWVKTLLINATQLLKLAKLAELVAAAIADISDVADI-----IKGTIGEVVEFTNAL 204
Db 236 ERAAAS-----AYLASFQSAFOGDVRAGVPGIDNLMRVKGDGSGGGL---AL 281
QY 205 NGLKELWDLKLTGWVT-GLFSRGS-----NLESEFFAG-----235
Db 282 -GMSDIAGSAGFTTHGLTFQGSAAAGGGGTAHAHVSAGAGDFGAVADNLGLDVGDLQTS 340
QY 236 VPGLTGATGSLGSOVTGLFGAAGLSASSG-----263
Db 341 TASVAGAGSTLSNASGLSGVGNVAGGAGGAMAAAPFMGMNGRHGPTAGQRFNSAGLNP 400
QY 264 ---LAHADSLASSASLAPALAGIGGSGFGGLPSL-AQVHAASRAQALRPADGPVGA 319
Db 401 NARTQASAMGPMGSMGLSGAGGAGVGGVGNRQALAPQGTGNPASAAGA 460
QY 320 QV-EGGSQLVSAQSGSQMGVPGV-----GMGMHPSSGASKGTTTKYSEGAAGTDEA 371
Db 461 QTPSGMNTGGVGRKASQSTARPMPMPMGAGGAGQPKNTGKVTVTSAVEQDANLAALIG 520
QY 372 ERAPV 376
Db 521 DRGPV 525

RESULT 7
Q06810 PRELIMINARY; PRT; 1329 AA.
AC Q06810
DT 01-JUL-1997 (TEMBLrel. 04, Created)
DT 01-JUL-1997 (TEMBLrel. 04, Last sequence update)
DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
DE Hypothetical protein Rv1450c.
GN Rv1450C OR MTCV493.04.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv.
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544 (1998).
DR EMBL; Z95844; CAB09271.1; -.
DR TuberculList; Rv1450c; -.
DR InterPro; IPR000084; PE_region.
DR InterPro; IPR002173; PFKB.
DR Pfam; PF00934; PE; 1.
DR ProDom; PD001223; PE_region; 1.
DR PROSITE; PS00583; PFKB_KINASES_1; 2.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 1329 AA; 107410 MW; 777A125F6DBAB234 CRC64;

Query Match      7.9%; Score 155; DB 16; Length 1329;
Best Local Similarity 26.3%; Pred. NO. 0.22;
Matches 79; Conservative 27; Mismatches 132; Indels 62; Gaps 13;

```

```
QY 110 EFVRPVAVDLTYIPVVGHALSAFAFPFCAGAMAVV---GGALAYLVVKTLLN----- 159
Db 9 ETVAALALDVARI---GSSIGANAA--AAGSTTSVLAAGADEVSAATLFGSHAREYQ 63
QY 160 --ATQLLKLLAKLAELVAAIAADIISDVADIKGTGLGEVWEFITNALN-----GLKEL 210
Db 64 AISTQVAAFHDFRFAQTLTSAAGSVYSAAEA-TNAAPLATLHNVLNALNAPTQALLGRPLI 123
QY 211 WDKL-----TGMVTGLFSRGNLSFFAGVPLTGCATSLQVTLFGAAGLSASSGLA 265
Db 123 GDAAGAPGTGGAGGAGGILWNGGAGSGAPQVGGAGG---AAGLFTGTGAGGAGGAG 179
QY 266 HADSLASSASLPALAGIGGSGFGGLPSLAQVHAASTRAQALPRADGPVCAAAEQVGQS 325
Db 180 AAGGAGSGGW--LLNGGVGGAGG-----QSL-----LGGATGGAGGNA 217
QY 326 QLVSAQSGQGMGPVGMGMHPSSGAS--KGTTTKYSEGAAGTDAERAPVEADAGG 383
Db 218 GLFGVGGTGGPGGPGGPGVGGTGGAGGLGGLTLYGAGGGGAGGP-----GPIGGVGGH 272

RESULT 8
QSVK17 PRELIMINARY; PRT; 1408 AA.
AC QSVK17;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE PE_PGRS family protein.
GN MT1497.1.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE007019; AAK45760.1; -.
DR TIGR; MT1497; -.
DR InterPro; IPR000084; PF0934; PE; 1.
DR Pfam; PF00934; PE; 1.
DR PRINTS; PR01223; EGGSHLL.
DR PRODOM; PD001223; PE region; 1.
SQ SEQUENCE 1408 AA; 113502 MW; EF472A34966AE6DC CRC64;

Query Match 7.7%; Score 152; DB 16; Length 1408;
Best Local Similarity 26.0%; Pred. No. 0.37;
Matches 78; Conservative 27; Mismatches 133; Indels 62; Gaps 13;

QY 110 EFVRPVAVDLTYIPVVGHALSAFAFPFCAGAMAVV---GGALAYLVVKTLLN----- 159
Db 10 ETVAALALDVARI---GSSIGANAA--AAGSTTSVLAAGADEVSAATLFGSHAREYQ 64
QY 160 --ATQLLKLLAKLAELVAAIAADIISDVADIKGTGLGEVWEFITNALN-----GLKEL 210
Db 65 AISTQVAAFHDFRFAQTLTSAAGSVYSAAEA-TNAAPLATLHNVLNALNAPTQALLGRPLI 123
QY 211 WDKL-----TGMVTGLFSRGNLSFFAGVPLTGCATSLQVTLFGAAGLSASSGLA 265
Db 124 GDAAGAPGTGGAGGAGGILWNGGAGSGAPQVGGAGG---AAGLFTGTGAGGAGGAG 180
QY 266 HADSLASSASLPALAGIGGSGFGGLPSLAQVHAASTRAQALPRADGPVCAAAEQVGQS 325

Query Match 7.7%; Score 152; DB 16; Length 1408;
Best Local Similarity 26.0%; Pred. No. 0.37;
Matches 78; Conservative 27; Mismatches 133; Indels 62; Gaps 13;

QY 110 EFVRPVAVDLTYIPVVGHALSAFAFPFCAGAMAVV---GGALAYLVVKTLLN----- 159
Db 10 ETVAALALDVARI---GSSIGANAA--AAGSTTSVLAAGADEVSAATLFGSHAREYQ 64
QY 160 --ATQLLKLLAKLAELVAAIAADIISDVADIKGTGLGEVWEFITNALN-----GLKEL 210
Db 65 AISTQVAAFHDFRFAQTLTSAAGSVYSAAEA-TNAAPLATLHNVLNALNAPTQALLGRPLI 123
QY 211 WDKL-----TGMVTGLFSRGNLSFFAGVPLTGCATSLQVTLFGAAGLSASSGLA 265
Db 124 GDAAGAPGTGGAGGAGGILWNGGAGSGAPQVGGAGG---AAGLFTGTGAGGAGGAG 180
QY 266 HADSLASSASLPALAGIGGSGFGGLPSLAQVHAASTRAQALPRADGPVCAAAEQVGQS 325
```

```
Db 181 AXGAGSGSGW--LLNGGVGGAGG-----QSL-----LGGATGGAGGNA 218
QY 326 QLVSAQSGQGMGPVGMGMHPSSGAS--KGTTTKYSEGAAGTDAERAPVEADAGG 383
Db 219 GLFGVGGTGGPGGPGGPGVGGTGGAGGLGGLTLYGAGGGGAGGP-----GPIGGVGGH 273

RESULT 9
QSVK15 PRELIMINARY; PRT; 738 AA.
AC QSVK15;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE PE_PGRS family protein.
GN MT1499.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE007019; AAK45763.1; -.
DR TIGR; MT1499; -.
DR InterPro; IPR002952; Eggshell.
DR InterPro; IPR000084; PE region.
DR Pfam; PF00934; PE; 1.
DR PRINTS; PR01228; EGGSHLL.
DR PRODOM; PD001223; PE region; 1.
SQ SEQUENCE 738 AA; 50853 MW; A6F4645310B9D0D0 CRC64;

Query Match 7.7%; Score 151; DB 16; Length 738;
Best Local Similarity 26.6%; Pred. No. 0.17;
Matches 80; Conservative 25; Mismatches 132; Indels 64; Gaps 13;

QY 110 EFVRPVAVDLTYIPVVGHALSAFAFPFCAGAMAVV---GGALAYLVVKTLLN----- 159
Db 10 ETVAALALDVARI---GSSIGANAA--AAGSTTSVLAAGADEVSAATLFGSHAREY 63
QY 160 --ATQLLKLLAKLAELVAAIAADIISDVADIKGTGLGEVWEFITNALN-----GLKE 209
Db 64 AISTQVAAFHDFRFAQTLTSAAGSVYSAAEA-TNAAPLATLHNVLNALNAPTQALLGRPL 122
QY 210 LWDKL-----TGMVTGLFSRGNLSFFAGVPLTGCATSLQVTLFGAAGLSASSGL 264
Db 123 GDAAGAPGTGGAGGAGGILWNGGAGSGAPQVGGAGG---AAGLFTGTGAGGAGGA 179
QY 265 AHADSLASSASLPALAGIGGSGFGGLPSLAQVHAASTRAQALPRADGPVCAAAEQVGQ 324
Db 180 AAGGAGSGSGW--LLNGGVGGAGG-----QSL-----LGGATGGAGGNA 217
QY 325 SOLVSAQSGQGMGPVGMGMHPSSGAS--KGTTTKYSEGAAGTDAERAPVEADAGG 382
Db 218 AGLFGVGGTGGPGGPGGPGVGGTGGAGGLGGLTLYGAGGGGAGGP-----GPIGGVGGH 272
QY 383 G 383
Db 273 G 273

RESULT 10
Q06808 PRELIMINARY; PRT; 741 AA.
ID Q06808
AC Q06808;
```

```

DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein RV1452c.
GN RV1452C OR MTCY493.02.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Emdolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL021942; CAA17449.1; -.
DR EMBL; AE006958; AAK44830.1; -.
DR TIGR; MT0607; -.
DR TubercuList; Rv0578c; -.
DR InterPro; IPR002952; Eggshell.
DR InterPro; IPR000209; Peptidase_S8.
DR InterPro; IPR000084; PE_region.
DR InterPro; IPR002173; PfkB.
DR Pfam; PF00934; PE; 1.
DR PRINTS; PR01228; EGGSHLL.
DR ProDom; PD001223; PE region; 1.
DR PROSITE; PS00583; PFkB KINASES 1; 2.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW COMPLETE proteome.
FT CONFLICT 363
FT CONFLICT 363
SQ SEQUENCE 1306 AA; 105964 MW; 843A30955FFA56B6 CRC64;

Query Match 7.6%; Score 150; DB 16; Length 1306;
Best Local Similarity 28.3%; Pred. No. 0.44;
Matches 60; Conservative 10; Mismatches 82; Indels 60; Gaps 8;

QY 216 GWVTGLFSRGWSNLESFFAGVPGLTGATSLGSOVTLFGAAGLSASSGLAHADSLASSAS 275
DB 185 GWLYGNGGAGG-----FGGA-GAVGNGGAGGTAGLFGVGGAGGAGNGIAGVTGTSAS 237

QY 276 LP-----ALACTGGSGFGGL-----PSLAQVHAASRQALRPAA 310
DB 238 TPQSGGTAGAGGTGGNGGAGGAGGVLMNGNGGAGGCGGCGGAGAGAGAGAHATNLGA 297

QY 311 DGPVGAARAEQGGSQLVSAQSGSQMGGP-----YGMGMH-----PSSG 350
DB 298 DG-----QAGGNGGNGGAGGTGGVGGPGGSHGHLGLGSHGAGGAGGCGGAGGAPGDG 350

QY 351 ASKGTITTKYSEGAAGAATEDAERAPVEADAGG 382
DB 351 NGATGTWGNLGG-AGGTGNGNGNPGAGGAGG 381

RESULT 12
O80116 PRELIMINARY; PRT; 916 AA.
ID O80116
AC O80116;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DE FGSR-family protein (PE FGSR family protein).
GN Rv0578C OR MT0607 OR MT039.16C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.

```

```

GN Bacteriophage b1l170.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
NCBI TaxID=63118;
FN [1]
RP SEQUENCE FROM N.A.
RA Crutz-Le Coq A.M., Ceseslin B., Commissaire J., Anba J.,
RA Kyriakidis S., Chopin M.C.;
RT "Sequence and organization of the lactococcal isometric b1l170 phage genome.";
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF009630; AAC37195.1; -
SQ SEQUENCE 916 AA; 96596 MW; 101058AE12E195F CRC64;

Query Match 7.6%; Score 149.5; DB 9; Length 916;
Best Local Similarity 21.8%; Pred. No. 0.29;
Matches 89; Conservative 58; Mismatches 144; Indels 117; Gaps 18;

Qy 1 MSRAFIIDPTTSA-----IDGLYDLLGIGIPNOGGILYSLEYFE--KALRE----- 45
Db 357 LSQALI--PTLEATPVNALIDKVDWKGAGAKIENVVKYFDOLFQLQONGAITQFS 414
Qy 46 -----LAAAPFG-DWLGSADKYAG-----KN-----RNVNPFQELADLRQLISLIH 89
Db 415 AIWNLKSAFGSVTGIIIGNLIKSPAGVDESTSKNSTSVENVAITSSLANKFADITKKIA 474
Qy 90 DQANAVQTRDILSGAKKGFSEVPAYDL-----TYTPVV--GHALSAAFPAP 136
Db 475 DFIOKISESKAMDAIKVALVALAGAFVMKVINGTIKAYETYNKIVEAGTTIOCAFNAI 534
Qy 137 FCAGAMAVGGALAYLVKTLINATQL---LKLAKLAELVAAA-----IA 179
Db 535 MAVNPFFVLIGIAAIVAGLVYFTQTGTGKAWASFVDFLKSANDGIVSFSGIGQWFA 594
Qy 180 DIISDVADIKGT-----LGEVWEPI-----TNALNGKLMDKLTGW 217
Db 595 DIWANGADVGAKEIWQGLVDWFIVGIQQIONINWGITTFTTWTVTWIGGIOSVWGVTGF 654
Qy 218 VTGTF-----SRGWSNLESFFAGVPGLTG----ATSGLSQVTLGPGA 255
Db 655 PSGIFDAVSSVTVSFSAIGGFASSANNVLVSMSAVAGFGGIFNAVSGV-VSSFSA 712
Qy 256 AGLSASGLAHADSLASAS-----LPALAGIGGG-SRGGGLPSLIA 295
Db 713 IGPSSAMGVVQSIWASVSGOFFSFGIFNSVSVSVSVFSAALGGFASNA 760

RESULT 13
Q8VU19 PRELIMINARY; PRT; 628 AA.
AC Q8VU19
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 23, Last annotation update)
DE PE PGRS family protein.
MT3476.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinomycetales; Actinomycetaceae;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
NCBI TaxID=1773;
FN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE007154; AAK47814.1; -
DR TIGR; MT3476; -

```

```

RA Bishai W.;
RT "whole genome comparison of Mycobacterium tuberculosis clinical and
RL laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL022004; CAA17639.1; -
DR EMBL; AB006974; AAK45096.1; -
DR TIGR; MT0834.1; -
DR TubercuList; Rv0833; -
KW Complete proteome.
FT CONFLICT 514 G -> A (IN REF. 2).
FT CONFLICT 584 S -> G (IN REF. 2).
SQ SEQUENCE 749 AA; 57448 MW; C3C91307F9A8A2EC CRC64;

Query Match 7.5%; Score 148.5; DB 16; Length 749;
Best Local Similarity 31.2%; Pred. No. 0.26;
Matches 53; Conservative 11; Mismatches 83; Indels 23; Gaps 4;

QY 220 GLFSRQNSNLESPFFACVPGLTGATSLQVTLFGAAGLSASSGLAHADSLASSASL PAL 279
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 27 GVGAGGAGGDSAVAGVIGGAGGAGGAALLFGAGGAGGAGSGSGSAAAGGAGGAGGGL 86
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 280 AGIGGGSGFGLPSLAQVHAASRQALRPADPVGCA--AAQVGGQSQOLNSAQSGMG 337
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 87 PASGGSGGFGGFASTGTGAGGT-----GGAGGLFASGGVGGTGGGAGSGGTGGVG 137
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 338 GPVGMGMHPSSGGA-----SKFTTKKYSAGAAAGTDAERAPVEADAGG 382
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 138 GTGGAGGLFASGGAGGAGSGGT-----GGAGGTGGAGGLFGAGGAGG 180
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 15
P95182 PRELIMINARY; PRT; 409 AA.
AC P95182;
DT 01-MAY-1997 (T-EMBLrel. 03, Created)
DT 01-AUG-1998 (T-EMBLrel. 07, Last sequence update)
DT 01-OCT-2002 (T-EMBLrel. 22, Last annotation update)
DE RV3144C (PPE family protein).
GN RV3144C OR MT3231 OR MTCY03A2.14.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaiia F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RL complete genome sequence.";
RL Nature 393:537-544 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umavam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RL laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z83867; CAB06293.1; -
DR EMBL; AE007137; AAK47571.1; -
DR TIGR; MT3231; -
DR TubercuList; Rv3144C; -

```

```

DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
KW Complete proteome.
FT CONFLICT 226 S -> G (IN REF. 2).
SQ SEQUENCE 409 AA; 39098 MW; 1D0C6483377855A2 CRC64;

Query Match 7.5%; Score 148; DB 16; Length 409;
Best Local Similarity 24.2%; Pred. No. 0.12;
Matches 94; Conservative 36; Mismatches 137; Indels 122; Gaps 15;

QY 12 SAIDGLYDLLIGIGIPNQGIGI---LYSSLYFEKALEELAAAPGCGWLGSAAADKYAGKNR 68
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 29 AAWDGLAEELGTAAQSFASVTAGLAQAWGCPAALAAVAAAAPYAGWLTAARAAQSGAG--- 85
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 69 NHVNFQELADLRQLISLIHQANAVQTTDILEGAKKGLFVRPVAVDLTYIPVVGHA 128
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 86 -----AAQARAVAS---IFEAAQAA-----TVLP----- 107
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 129 LSAFAQAPFCAGANAVVGALAYLVVKTLLINATQLLKLAKLAELVAAAADIISVDADI 188
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 108 -----AAVAANRDFAVQLV-----MTNLFQONAPLIAA----- 136
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 189 IKGTGLGEVWEFITNALNGLKELWDLTGWVTGLFSRGSWNLESFFAGV-PGLTGATSGLS 247
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 137 -EGVYEMWAADVAAMSGYSGASATAQV-----PWASLLQRFPLGAGATGATGGS 190
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 248 QVTGL-----FGAAGLSAS-SGLAHADSLASSASL-----PALAGIGGSGGFG 289
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 191 VGTGATGGSVGTGGGESVGTGATASGGGVGVGSVASAGLAAGDPAHGSVGGQNFQFG 250
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 290 GLPSLAQVHAASRQALRPADGPGVGAAGAEQVGQSQOLNSAQSGCMGPGVGMGHPSS 349
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 251 GDVGAGDVVASSATSA-----HAGVSPGFTGAPLAALALQKARGG-----T 293
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 350 GASKGTTTKKYSAGAAAGTDAERAPVEA 378
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 294 NSAPGTATE-----SARAPEPAASAPPEA 317
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Search completed: November 5, 2003, 17:02:38
Job time : 59 secs